Towards Intelligent Immersive Healthcare: A Systematic Review on the Integration of AI with AR/VR in Medical Applications

FIT 5128 Master Thesis S1 2026

Author: Xiumiao Zhang

Student ID: 35294639

Supervisors: Jiazhou 'Joe' Liu, Tim Dwyer, Yidan Zhang, Agnes Haryanto

Word Count: 7850

Content

Part 1

General Literature Review / Proposal
(Resubmit)

Monash University

Faculty of Information Technology

Master of YOUR DEGREE

Literature Review

Towards Intelligent Immersive Healthcare: A Systematic Review on the Integration of AI with AR/VR in Medical Applications

Author:
Xiumiao Zhang

Student ID:
35294639

Supervisors:
Jiazhou 'Joe' Liu
Tim Dwyer
Yidan Zhang

Semester 18, 2025
Fit 5126

Contents

1 Introduction

With the continuous development of information technology, immersive technologies (Immersive Technologies), which are a general term for interactive media such as virtual reality (VR), augmented reality (AR), and mixed reality (MR), embed digital information into human perception and interaction processes, allowing users to obtain immersive experiences in real or virtual environments[1, 2]. These technologies are gradually entering the medical field and have shown unique advantages in surgical training, rehabilitation, and psychological interventions[3]. For example, VR as an immersive technology can provide a safe and controlled training environment in medical education, significantly improving the knowledge, cognitive skills, and practical abilities of medical students and health professionals compared with traditional or other digital education methods[4].

With the growing maturity of these two technologies, some studies have found that although AI can provide relatively accurate diagnoses and predictions, it lacks interpretability, which creates barriers for its application in medicine[5]. On the other hand, the functions of immersive technologies still focus mainly on visual support, lacking intelligent prediction and real-time feedback mechanisms, which limits their potential in complex medical tasks. Therefore, researchers and industry have begun to explore the potential of combining AI with VR/AR, with healthcare being one of the main areas of focus[6]. This combination means that AI provides intelligent computing power for immersive technologies, while immersive technologies provide visual support for AI. The two influence and support each other, rather than working independently in different systems. Some research results have shown that the combination of AI and VR/AR may create a "synergistic effect." For example, Pelosi et al embedded reinforcement learning (AI) into an immersive VR rehabilitation game, dynamically adjusting training tasks based on patients' real-time movements, and confirmed that this method can achieve personalized and adaptive rehabilitation programs and improve training outcomes[7].

However, there is no systematic perspective to answer the current situation of the integration of AI and VR/AR in the medical field. From the existing literature, most review studies focus on the application of a single technology. For example, Sharma et al conducted a scoping review of studies on the implementation of artificial intelligence in healthcare practice from 2011 to 2022, but did not include immersive technologies[8]. On the other hand, Ghaednia et al, Selvan et al, and Mergen et al carried out systematic reviews of AR/VR in healthcare from perspectives such as spine surgery, vision testing, and medical training, but without analysis of AI support[9, 10, 11]. This separation of the two technologies has led to an insufficient understanding of how "intelligent + immersive" systems can be integrated.

Recent studies have started exploring the potential of integrating AI and immersive technologies, but they encounter clear limitations. One group of studies focuses on specific settings. For example, Chance explored how AI and VR can support medical education[12], while Tabassum et al reviewed how AI and extended reality can be integrated in mental health interventions, assessing their potential to improve outcomes and patient engagement[13]. Additionally, Rudnicka et al conducted a systematic review of AI and XR applications in the cardiac field[14]. However, these studies often have a narrow scope, and some do not cover a wide range of immersive technologies (VR/AR/MR). Building on these limited attempts, Wu et al conducted the first systematic review specifically on AI-powered VR systems in medical settings[15]. This study provided a comprehensive overview of AI and VR applications in medicine, categorising them into three groups: visualisation enhancement, VR-related medical data processing, and VR-assisted interventions, helping readers understand the field from a broad perspective. However, it did not include AR studies and lacked a strict search and statistical process, which may have resulted in missing papers and selection bias.

In summary, no study has yet systematically summarised and analysed how artificial intelligence (AI) and VR/AR are actually being applied in real-world practice. In other words, whether these AI + VR/AR technologies can truly enter medical practice and pass the strict tests of evidence-based medicine for safety, effectiveness, and patient acceptance remains a pressing question that urgently needs to be answered by systematic research.

This gap has shaped the primary focus of this study, which aims to conduct a systematic literature review to comprehensively examine and analyse the current state and trends of integrating AI with virtual and VR/AR in clinical trials. This strong emphasis on clinical applications ensures that our review remains highly relevant to guiding medical practice. The main research question guiding this study is: "Where and how is AI+AR/VR having or trending towards real-world impact in a clinical setting?"

To address these gaps, this review is structured around the main application areas for integrating AI and VR/AR: 2.1 Advances in Immersive Technologies in Medicine. 2.2 Advances in Artificial Intelligence in Medicine. 2.3 Integration of Artificial Intelligence and Immersive Technologies in Medicine.

2 Substantive Literature Review

This section reviews the progress of artificial intelligence and immersive technologies in medicine, with a focus on how they are used together in clinical settings. The review looks at three areas: first, the development of AI in medicine; second, the current use of immersive technologies in medicine; and third, the research progress on their integration. The goal is to clearly summarise the existing studies and build a base for finding research gaps.

2.1 Advances in Immersive Technologies in Medicine

2.1.1 Current Applications of Immersive Technology in Medicine

Over the past thirty years, immersive technologies have moved forward quickly in medicine thanks to their sense of presence, interactivity, and 3D visualisation. Since around 2010, when consumer VR headsets began to reach the market, improvements in display and tracking have lowered the barriers for research and use[16]. This has driven the rapid growth of VR in health care, with applications now extending to surgical training and intraoperative navigation, rehabilitation, mental health, and pain management.

In surgery, immersive technology has shown clear value. VR is mainly used for preoperative training, offering surgeons realistic simulation. Randomised controlled trials show that VR training reduces intraoperative complications and supports transfer of skills to clinical tasks[17, 18]. These findings establish VR in surgical education. AR, in contrast, is focused on intraoperative use, especially navigation. Studies show that AR systems improve planning and accuracy in neurosurgery and spine surgery[18, 19]. Clinical analyses report that AR with 3D imaging can shorten operation time, reduce incision length and improve patient recovery[20]. AR has moved from early fixed displays to head-mounted displays and smart glasses. Compared with traditional devices, head-mounted systems such as Augmedics xvision improve flexibility and accuracy, while portable devices such as Google Glass show potential in teaching, consultation and workflow efficiency[21]. However, challenges remain in battery life, resolution and data privacy.

In rehabilitation, adherence often decides success. Traditional exercises are repetitive and boring, leading to dropout. VR, with its immersion and engaging nature, offers new ways to improve participation. A randomised controlled trial by Ase et al[22]. showed that home-based VR training with the RAPAEL smart glove not only improved upper limb function in chronic stroke patients (FMA-UE, MAL), but all participants also completed the programme. This shows VR's advantage in adherence and usability[22]. Unlike VR, which emphasises engagement, AR improves accuracy and detailed management. A trial by Shim et al included 115 patients after rotator cuff repair[23]. Results showed that the AR-based digital rehabilitation system (UINCARE Home+) led to greater improvement in shoulder function (SST) at 12 weeks compared with traditional rehabilitation, with better trends in patient-reported outcomes such as SPADI, DASH and EQ-5D-5L. No significant differences were seen in pain, range of motion or strength, but AR's real-time feedback helped patients optimise movement and support recovery[23].

In mental health, VR is most established in exposure therapy [8]. Its key advantage is recreating fear or anxiety in safe and repeatable settings. A meta-analysis by Deng et al confirmed that VR exposure therapy reduces anxiety and PTSD symptoms[24]. Other studies tested wider access and automation. Freeman et al ran a randomised controlled trial where patients with fear of heights used an immersive VR system with a virtual coach. Within two weeks, symptoms improved significantly and lasted at follow-up[25]. This shows VR can reproduce traditional exposure therapy and also overcome therapist shortages, offering scalable mental health care.

In pain management, VR has been widely studied as a non-drug approach. Many trials confirm strong effects in acute pain relief. The early "SnowWorld" study Hoffman et al used distraction to reduce pain in burn patients during wound care, setting the foundation for VR pain relief[26]. Wang et al later showed that short preoperative VR exposure reduced pain after laparoscopic gynaecological surgery while improving anxiety, sleep and recovery, suggesting dual benefits in perioperative care: pain relief and psychological regulation[27]. In chronic pain, evidence is limited, but studies suggest VR with cognitive behavioural therapy can reduce pain catastrophising and support longer-term improvements[28].

Taken together, immersive technology shows the feature of "one technology, different logics" across fields. In surgery the focus is precision, in rehabilitation adherence, in mental health safety, and in pain management distraction. These differences suggest that the value of immersive technology depends on context and goals, rather than being universal Representative studies and results are summarised in Table 1.

Table 1: Summary of Studies on Immersive Technologies in Medicine

Application Area	Author(s)	Year	Technology Type	Device Used	Specific Application
Surgery	Raison et al.	2021	Virtual Reality (VR)	VR surgical simulator	Surgery training
	Seymour et al.	2002	Virtual Reality (VR)	MIST VR simulator (Mentice AB, Sweden) with laparoscopic interface (Immersion Corp)	Surgery training
	Ivan et al.	2021	Augmented Reality (AR)	AR HMD (head-mounted display)	Cranial neurosurgery (incision planning)
	Shi et al.	2025	AR + 3D-CT	AR navigation with 3D-CT reconstruction	Microvascular decompression (hemifacial spasm)
	Molina et al.	2021	Augmented Reality (AR)	AR-mediated stereotactic navigation	Lumbar spondylectomy (oncologic spine surgery)
	Wei et al.	2018	Wearable AR (Google Glass)	Google Glass	Surgical assistance / intraoperative guidance
Rehabilitation	Ase et al.	2025	Virtual Reality (VR)	Home-based VR rehabilitation system	Chronic stroke – upper extremity rehabilitation
	Shim et al.	2023	Digital healthcare system (VR/AR-based)	Mobile + wearable healthcare platform	Postoperative rehabilitation after rotator cuff repair
Mental Health	Freeman et al.	2017	Virtual Reality (VR)	Immersive VR platforms (various)	Mental health disorders (assessment & therapy)
	Deng et al.	2019	Virtual Reality (VR)	Immersive VR systems	PTSD – VR exposure therapy
	Freeman et al.	2018	Virtual Reality (VR)	Automated immersive VR therapy system	Specific phobia (fear of heights)
Pain Management	Hoffman et al.	2019	Immersive Virtual Reality (VR)	VR analgesia system	Acute pain management in pediatric/adolescent burn patients (Latin American)
	Wang et al.	2024	Virtual Reality (VR)	Oculus Quest 2 (Meta)	Acute postoperative pain (gynecological laparoscopic surgery)
	Carvajal-Parodi et al.	2025	Virtual Reality (VR)	Immersive & non-immersive VR interventions	Chronic pain – pain catastrophizing

2.1.2 Challenges of Immersive Technology in Medicine

Although immersive technologies in medicine, such as virtual and augmented reality, have grown rapidly in recent years and shown clear benefits across different areas, their use in everyday clinical practice still faces challenges. One major issue is the lack of real-time feedback and interaction. Whether in surgical navigation or rehabilitation training, most systems only provide visual support. They do not offer haptic or force feedback, which limits their ability to fully reproduce complex medical tasks. Another challenge is the lack of smart prediction and personalisation. Many immersive programs still use standard templates and do not adjust based on patient imaging, physiological data, or behavioural responses. This makes precise and tailored interventions difficult. A further problem is the heavy reliance on manual work. Clinicians are often required to supervise training, record progress, and analyse results, which increases their workload and reduces efficiency.

Because of these issues, it is clear that while immersive technologies have delivered promising results, their real value in medicine will depend on smarter and more adaptive support. The fast progress of artificial intelligence (AI) offers an important opportunity to fill this gap. AI can provide real-time feedback, enable personalised interventions, and reduce repetitive work by automating data capture and analysis. It can also support complex clinical decisions, such as tumour margin recognition or surgical route planning. The following section will review advances in artificial intelligence in medicine, with a focus on how AI can work together with immersive technologies to drive the future of smart healthcare.

2.2 Advances in Artificial Intelligence in Medicine

2.2.1 Current Applications of Artificial Intelligence in Medicine

Artificial intelligence (AI) has been explored in medicine for several decades. The earliest work can be traced back to the 1970s with rule-based expert systems, such as MYCIN, which was designed for infectious disease diagnosis and antibiotic advice[29]. However, these systems relied on rules set by humans and struggled to adapt to complex and changing clinical situations, so they were not widely adopted[30]. With the growth of computing power and data, especially the breakthroughs in deep learning in the 2010s, AI achieved major progress in medical imaging analysis, pathology recognition, and disease prediction[31]. Today, the use of AI in medicine has gradually moved from theory into clinical practice, playing a role in key stages of the healthcare process. Current studies show that the choice of AI models depends more on the application goal (such as diagnosis or prognosis/prediction) rather than on the medical field itself[29]. Researchers have also found that most applications of AI focus on diagnosis (44.4%), followed by prognosis/prediction (13.9%) and screening (9.3%). Based on this, this paper will analyse the current use of AI models in different medical fields through the lens of application goals[29]. As shown in the Table 2 below.

Table 2: Summary of Studies on Artificial Intelligence in Medicine

First Author	Year	AI Model Used	Device / Data	Medical Domain	Application Goal
van Ginneken	2015	OverFeat (pretrained CNN features + linear SVM)	Chest CT (LIDC dataset, 865 cases)	Radiology / Lung Cancer	Diagnosis / Screening
Gulshan	2016	Deep Convolutional Neural Network (CNN)	Retinal fundus images (EyePACS, Messidor-2)	Ophthalmology / Diabetic Retinopathy	Diagnosis / Screening
McKinney	2020	Deep Learning AI system	Mammography (UK + US screening data)	Radiology / Breast Cancer	Diagnosis / Screening
Kim	2024	Deep Learning software (AI-assisted CTA analysis)	Brain CTA (595 ischemic stroke patients)	Neuroimaging / Stroke	Diagnosis / Screening
Ardila	2019	End-to-end 3D CNN model	Low-dose chest CT (NLST + clinical validation)	Radiology / Lung Cancer	Diagnosis / Screening
Walsh	2017	Machine Learning algorithms (ensemble across multiple models)	Electronic Health Records (5,167 patients)	Psychiatry / Psychology	Prognosis / Prediction
Nguyen	2019	Deep Neural Network (DNN)	Surgical motion sensors (IMU data + JIGSAWS dataset)	Surgery / Education	Other (classification of surgical skill levels)
Khalid	2020	Deep Learning models (video analysis)	Surgical video clips (103 clips, JHU dataset)	Surgery / Performance Evaluation	Other (surgical action recognition & skill assessment)
Debnath	2022	Review of computer vision approaches (pose estimation, activity recognition, CV-based comparison)	Kinect, RGB/Depth cameras, other CV systems	Rehabilitation Medicine / Motor Disorders	Other (rehabilitation monitoring & assessment review)

In diagnosis and screening, medical imaging is without doubt the most breakthrough area for AI. Radiology and oncology are at the front line of its use. Convolutional neural networks (CNNs) are widely applied in image recognition, from lung nodule detection, retinal disease screening, and breast cancer screening to stroke image diagnosis. In many of these tasks, AI has reached or even passed the performance of radiologists across different fields [32, 33, 34, 35].

In medical , imaging-based prediction is still the most developed and widely studied direction. Ardila et al built a three-dimensional deep learning model that predicts lung cancer risk on low-dose CT scans[36]. Its accuracy was already higher than that of radiologists when no prior images were available, and it reached a comparable level when full image sequences were provided. This kind of work shows the strengths of AI in structured medical imaging, especially for large-scale screening and early detection. By contrast, prediction in mental health relies more on unstructured data and time-series information. Walsh et al used machine learning methods to analyse the electronic health records of more than 5,000 patients and developed a model to predict suicide attempts[37]. The model showed strong performance (AUC = 0.84, precision = 0.79, recall = 0.95), with prediction power increasing sharply in the seven days before a suicide attempt. These results highlight the unique value of AI in capturing the dynamic features of mental health crises and suggest possibilities for early intervention in this field.

Beyond diagnosis, prognosis/prediction and screening, AI is also being used in clinical process and functional assessment. Here the focus is not on whether a disease is present, but on making operations and motor functions objectively measurable to support medical training and personalised rehabilitation.

In surgery, Nguyen et al integrated deep neural networks with tool and hand motion signals to accurately distinguish skill levels in open and robot-assisted surgical tasks[38]. On the JIGSAWS dataset for suturing, needle passing and knot tying, their system reached or exceeded 95% recognition performance, showing strong potential as an alternative to traditional expert-based scoring [38]. Khalid et al built a deep learning model from intraoperative videos that could recognise specific surgical actions and quantify operator performance (average precision 0.97, recall 0.98; skill assessment precision/recall 0.77/0.78)[39]. This provided practical and objective indicators for training feedback and workflow improvement [39].

In rehabilitation, the review by Debnath et al showed that computer vision and pose estimation/action recognition can deliver automated and standardised assessment in gait analysis, upper limb function and remote rehabilitation, reducing reliance on manual observation. At the same time, the review noted limits in cross-setting generalisation, privacy concerns and lack of standardised metrics[40]. It suggested that future progress needs multi-modal fusion (video + kinematics + wearables), alignment with clinical scales for interpretability, and integration with immersive technologies such as AR/VR. These steps would help turn algorithmic accuracy into measurable benefits for training quality and functional recovery outcomes (Debnath et al., 2022).

2.2.2 Challenges of Artificial Intelligence in Medicine

Although artificial intelligence (AI) has shown great potential in medicine, its use still faces many challenges. A key problem is that many AI models act as a "black box". They can provide predictions with high accuracy, but the reasoning behind them is not transparent. In clinical work, this lack of interpretability can cause trust issues and make it difficult for doctors to understand or explain AI-based decisions to patients [5]. In addition, the development of AI is still limited. Even high-performing models are not always correct, and in sensitive areas such as diagnosis or treatment planning, errors can have serious consequences [41, 42]. Because of this, it is important to keep a "human in the loop" in medical applications. Instead of letting algorithms work alone, medical experts should be involved to fine-tune AI outputs using their knowledge and clinical experience. This human-AI collaboration can make use of the efficiency of AI while adding expert oversight to improve the safety and reliability of decisions[42].

In this context, improving the interpretability and usability of AI is especially important. In recent years, new immersive technologies, such as augmented reality (AR) and virtual reality (VR), have provided more intuitive ways to visualise complex AI predictions. These tools can help doctors better understand and apply model outputs, supporting the smoother use of AI in clinical practice.

2.3 Integration of Artificial Intelligence and Immersive Technologies in Medicine

AI integrated with immersive technologies (AI+XR) is used most heavily in medicine. The systematic review by Reiners et al shows that most AI+XR studies focus on medical and health settings[43]. This reflects the strong clinical need for intelligent visual tools and suggests that medicine is becoming the key testbed for bringing AI+XR into practice. It is therefore necessary to review the integrated use of AI and immersive technologies in medicine, not only to summarise current results but also to guide future research and practice.

As noted earlier, AI and immersive technologies have each made progress in medicine, but both have limits when used alone. XR often lacks real-time feedback and personalisation, so it struggles with complex clinical tasks. AI has strong prediction and recognition ability, but without clear visualisation it is hard for clinicians to understand and adopt. Their combination is naturally complementary: AI adds intelligence and adaptivity to immersive systems, while XR turns complex AI outputs into visual and interactive forms that are easier to use at the bedside. This mutual reinforcement can not only cover each other's weaknesses, but may also create a "1+1¿2" synergy. At present, AI+XR applications cluster in surgical navigation and training, rehabilitation, mental and neurological interventions, and newer areas of clinical support and medical education.

Surgery and intraoperative navigation are the most direct use cases. Chiou et al used deep learning for automatic CT segmentation and linked it with AR navigation on HoloLens, providing real-time 3D guidance and improving accuracy[44]. They also noted hardware differences: tablet-based systems had better spatial accuracy than head-mounted devices, showing that XR's benefits still depend on device stability. Likewise, Pierre et al integrated AI image processing with AR display for vertebroplasty navigation, cutting radiation exposure and improving path planning[45]. Together, these studies show that AI delivers intelligent segmentation and recognition, while AR presents results in an intuitive way to support safer, more efficient surgery. In this field, most AI+XR integrations follow a one-way interaction model: AI outputs flow into the AR/VR display, but user actions in XR do not feed back to change the AI.

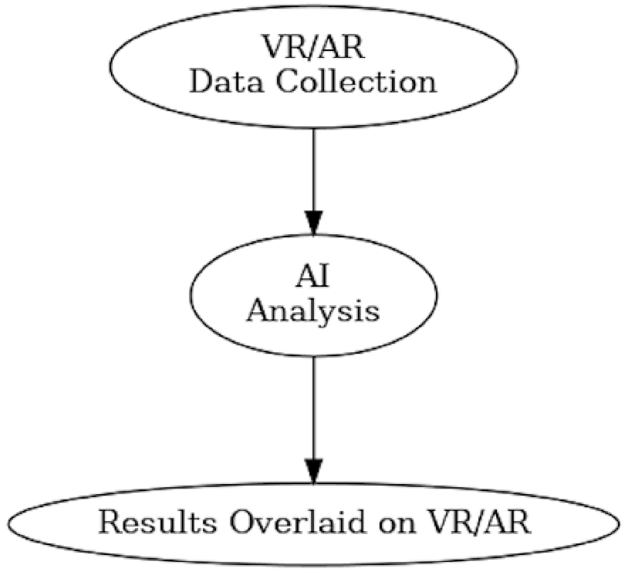

Figure 1: one-way integration

In rehabilitation, AI+XR often forms a closed-loop, real-time interaction. VR/AR provides an adaptive training environment; patients perform tasks; AI analyses behaviour and generates feedback; the VR/AR scene then updates accordingly. Pelosi et al applied reinforcement learning to adjust task difficulty on the fly for upper-limb rehab in VR, creating a "task–feedback–retraining" loop and improving personalisation[7]. Wang et al used an AI coaching agent in an adolescent obesity program[46]; the AI adjusted exercise intensity and pacing from live data to boost effectiveness. Unlike surgery, rehab emphasises adaptive feedback, where AI's immediate reading of behaviour changes the virtual environment.

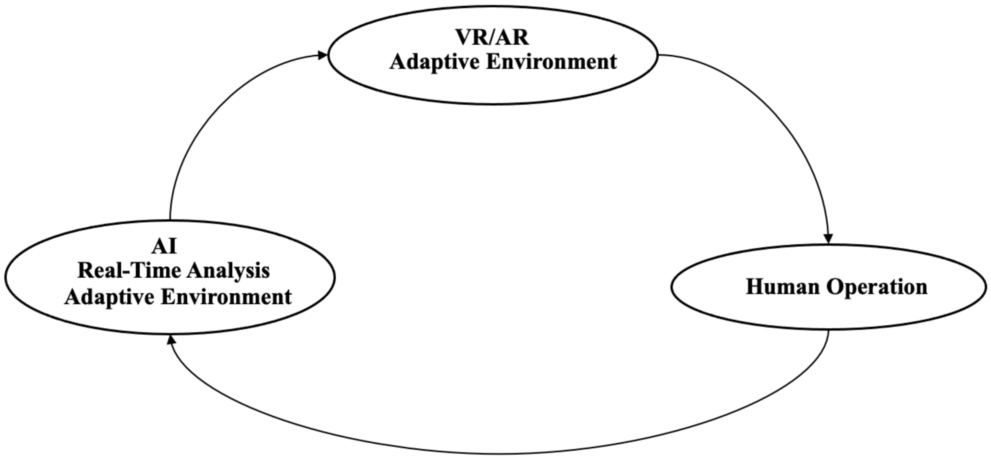

Figure 2: closed-loop integration

In medical education and training, Fazlollahi et al showed that an AI teaching assistant in a VR surgical simulator can provide voice and visual feedback based on real-time performance, improving learning outcomes[47]. A randomised controlled trial further found that AI plus human expert guidance outperformed either alone [48]. This suggests that in education the aim is not to replace experts, but to augment them. Here AI acts more as an "intelligent tutor", giving prompts and scoring feedback, rather than adapting the environment itself.

For mental and neurological interventions, AI+XR often uses a data-driven offline integration. VR/AR collects behaviour or image data; AI then performs diagnosis, classification, or prediction; results are not fed back to XR in real time. Tsai et al targeted early neurocognitive detection: patients completed tasks in a VR supermarket, actions were recorded, and machine learning classified them as MCI, Alzheimer's, or healthy[49]. Oh et al studied ADHD: patients did attention and impulse-control tasks in VR, and AI models analysed the data to aid diagnosis and rating[50]. Hudon et al examined schizophrenia with Avatar Therapy: transcripts of patient–avatar dialogues were used for machine learning to classify themes and predict outcomes[51]. Duan et al worked on depression prognosis: VR social interaction tasks revealed self-blame patterns, and AI predicted symptoms four months later[52]. These studies show that AI can extract complex features from unstructured behaviour and language, while VR provides a standardised, controllable setting that enables prediction and tailored intervention. However, a multi-centre randomised controlled trial by Lendaro et al indicates that AI does not always add value: in phantom-limb pain, an AI+AR intervention based on myoelectric decoding (PME) had similar efficacy to a VR imagery-based intervention (PMI)[53]. This invites caution: AI and XR do not deliver extra benefit in every setting, and future work should identify the conditions under which their combination is truly more effective.

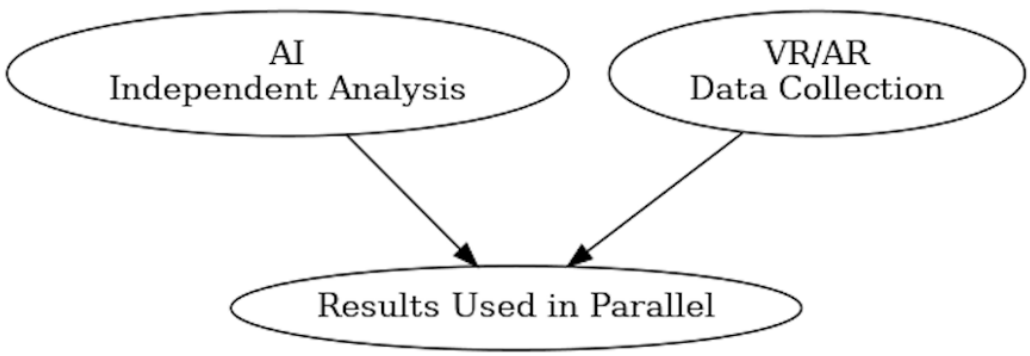

Figure 3: data-driven offline integration

3 Summary of the State of the Art

From the above review, it is clear that current applications of artificial intelligence with immersive technologies (AI+XR) in medicine are concentrated in surgical navigation and training, rehabilitation, mental and neurological interventions, medical education, and emerging forms of clinical support. These uses show the pair's complementarity: AI provides real-time analysis and personalisation, while XR turns complex AI outputs into visual, interactive displays. However, the literature also shows clear limits. Most studies remain at lab tests or small feasibility trials, with few multi-centre, large-sample evaluations in real-world clinical settings. In short, while current evidence shows promise, whether AI+XR can meet the rigour of evidence-based medicine is still an open question.

This underscores the need for a systematic literature review. On one hand, findings are spread across diverse use cases and technical pathways, with little synthesis of overall trends. On the other, existing reviews often focus on a single technology (e.g., AI in imaging, or VR/AR in education and rehabilitation) rather than their combination. Summarising the joint use of AI and XR in medicine can clarify which scenarios truly benefit from synergy, and also when AI may not be necessary.

In medical research, clinical trials are usually seen as the key form of validation. Trials test safety and effectiveness under controlled conditions and also show feasibility and patient acceptance in practice. Compared with lab or simulated studies, trial results carry more weight with clinicians and are more likely to move AI+XR from "proof of concept" to clinical adoption. Accordingly, this study will conduct a systematic review focused on clinical trials of AI+XR in medicine, aiming to answer the core question: Where and how is AI+AR/VR having or trending towards real-world impact in a clinical setting?

4 Research Project Plan

The purpose of this literature review is to provide a preliminary overview of the current state and development of artificial intelligence (AI) and immersive technologies (XR, including VR and AR) in the medical field. By examining the independent progress of AI and XR in medicine, as well as their combined applications in areas such as surgical navigation, rehabilitation medicine, medical education, and mental health interventions, this review not only summarizes their advantages and limitations but also highlights their complementarity and potential synergies. At the same time, by analyzing existing research outcomes, this study identifies methodological shortcomings and gaps, laying the theoretical and practical foundation for a more in-depth systematic review—particularly one focused on clinical trial evidence. This chapter will outline the research plan in detail, including research questions and objectives, study design and methodology, data collection strategies, experimental aims, and ethical and privacy considerations, ultimately providing a research framework for a systematic evaluation of AI+XR in the medical domain.

4.1 Research Question and Aims

The main research question of this study is: Where and how is AI+AR/VR having or trending towards real-world impact in a clinical setting?

To solve this question, this study is mainly divided into the following sub-questions:

Question 1 (Application Areas):

What are the current clinical application areas where AI and AR/VR technologies are used together?

Question 2 (Technology):

What types of AI methods and AR/VR platforms or devices are commonly used in the existing studies? How do they work together?

Question 3 (Implementation and Evaluation):

How effective are these AI + AR/VR applications in real medical practice?

Question 4 (Trends and Gaps): What are the current research trends in the integration of AI and VR/AR technologies in healthcare, and what are the key limitations or gaps in existing studies?

4.2 Research Design and Method

4.2.1 Research Design

This study adopts a Systematic Literature Review (SLR) design to ensure comprehensiveness, transparency, and reproducibility. Unlike a general narrative review, it will strictly follow the PRISMA framework, covering database search, inclusion and exclusion criteria, data extraction, and quality assessment. This approach guarantees the credibility and academic value of the review findings, while also providing a structured methodological framework and practical insights for future research and clinical implementation.

4.2.2 Research Method

The methodology of this study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework and consists of the following steps:

Database selection: PubMed, IEEE Xplore, Scopus, and Web of Science.

Time frame: 2020–2025, covering the period of rapid development of AI and XR technologies in medicine over the past five years.

Search keywords: ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning") AND ("Virtual Reality" OR "Augmented Reality" OR "Mixed Reality" OR "Extended Reality") AND ("Clinical Trial" OR "Randomized Controlled Trial" OR "Pilot Study").

Table 3: Inclusion/Exclusion Criteria

Inclusion Criteria	**Exclusion Criteria**
Integration of AI with AR/VR/MR	AI and AR/VR/MR used independently, with no interaction/integration
Medical/health domain	Non-medical/health domain
Within the scope of clinical trial research test clinical tial condition	Non-clinical-trial study
Human participants/patients（alive human）	Animal/cadaver/phantom/simulated subjects
Empirical study	Literature review, commentary, or meta-analysis,design paper(no evaluation)
Written in English	Written in other languages
peer-reviewed	Non–peer-reviewed

The inclusion criteria require that the literature must involve the combined application of AI and AR/VR/MR, rather than the use of either technology alone, with a certain level of interaction between the two. The research must be within the medical or health domain and fall under the scope of clinical trials, involving living human participants or patients rather than animals or virtual simulations. The studies must be empirical in nature, written in English, and published in peer-reviewed journals.

The exclusion criteria specify that studies will be excluded if they involve only the independent use of AI or XR without integration; if the research is outside the medical or health domain; if it does not fall within the scope of clinical trials; if the subjects are animals, cadavers, prostheses, or simulated participants; if the work is limited to literature reviews, commentaries, or methodological papers without empirical data; or if the studies are not written in English or have not undergone peer review.

In short, these criteria ensure that the review focuses on empirical studies of AI+XR integration conducted in real clinical settings with human participants, thereby maximizing the clinical value and reliability of the findings.

4.3 Data Collection

Data Extraction and Recording For the final set of included studies, a structured data extraction form will be developed. The main variables will include:

1.Basic information: author, year, journal, country/region of study

2. Technical characteristics: type of AI method, XR platform or device

3. Integration approach: the specific mode of combining AI with immersive technologies

4. Application scenarios: surgical navigation and training, rehabilitation, medical education, mental and neurological disorders, etc.

5. Study design: type of trial (e.g., RCT, prospective study), sample size, follow-up duration

6. Primary outcomes: accuracy, efficiency, adherence, clinical effects, etc.

7. Limitations and future directions as discussed by the authors. Through this process of data collection and management, the completeness and reliability of the dataset will be ensured, providing a solid foundation for subsequent systematic analysis and comparison.

4.4 Experimental Goals

The experimental objectives of this study are closely aligned with the core research questions—namely, where and how the integration of artificial intelligence (AI) and augmented/virtual reality (AR/VR) is currently generating, or tending to generate, real-world impact in clinical settings. The specific objectives are as follows:

1. Organize clinical application areas

Systematically identify and categorize medical application scenarios of AI combined with AR/VR—such as surgical navigation, rehabilitation, medical education, and interventions for neurological and psychiatric disorders—in order to clarify their relevance and scope in clinical practice.

2. Characterize modes of technological integration

Summarize and analyze the specific AI methods used in existing studies (e.g., deep learning, reinforcement learning, natural language processing) together with XR platforms (e.g., VR headsets, AR-based surgical navigation systems), and explore how these technologies interact to form integrated solutions.

3. Assess clinical implementation and effectiveness

Synthesize empirical findings from clinical trials, with particular attention to the effectiveness of AI+XR applications in real-world medical practice. Key indicators include diagnostic accuracy, procedural efficiency, patient adherence, clinical outcomes, and technological maturity.

4. Identify trends and research gaps

Summarize the latest development trends in AI+XR integration and highlight the major shortcomings of current research. By achieving these objectives, this study will not only consolidate the current evidence base but also provide directional insights for the future development of AI+XR in medicine. More importantly, the findings will help researchers, clinicians, and policymakers understand where these technologies have the greatest impact, where their limitations lie, and why focusing on clinical trials is essential for translating technological innovation into real medical practice.

4.5 Ethics and Data Privacy

As a systematic review, this study does not directly involve patient data and therefore carries minimal ethical risk. However, when evaluating the included studies, special attention will be given to the following aspects:

1. Whether the clinical trials obtained ethical approval (IRB/Ethics Committee);

2. Whether data usage complies with privacy regulations such as GDPR and HIPAA;

3. The interpretability and safety of AI algorithms applied to patient data.

4.6 Project schedule

The expected duration of the whole research project is 11 months, starting from July 2025 and ending in June 2026. Literature review and research protocol design will be completed in the first to second months, followed by the search strategy and database searching in the second to third months. Screening of abstracts and full texts as well as data extraction will be conducted from the fourth to the seventh months.

Statistical analysis and report writing will be carried out in the eighth to the tenth months. The final month will be dedicated to collating the results, preparing submissions to academic conferences or journals, and finalizing the project.

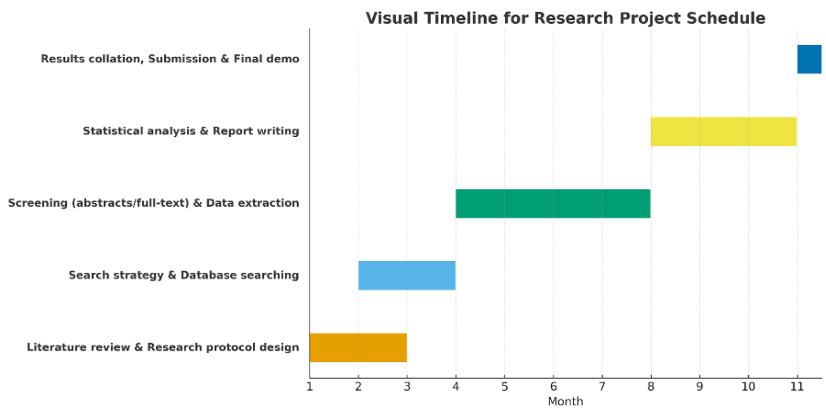

Figure 4: Prject Schedule Gantt chart

5 Conclusion

This review looked at the separate progress of immersive technologies and artificial intelligence in medicine. It found that immersive technologies still face limits such as the lack of real-time multi-modal feedback, limited personalisation, and heavy reliance on manual supervision. Artificial intelligence also has problems, including the "black box" issue, a lack of clear visual support, error risks in sensitive areas, and the need to keep a "human in the loop". Building on these insights, this study then examined evidence of how the two are used together in surgery, rehabilitation, mental and neurological health, and medical education. The results show that AI and XR can complement each other—AI provides prediction, recognition and adaptive support, while XR improves visualisation, guidance and engagement. At the same time, current studies are still limited by small sample sizes, single-centre designs, a lack of real-time AI use, and uneven strength of evidence across fields. To address these issues, this thesis sets out a PRISMA-based systematic review plan centred on clinical trials. It defines the research questions and aims, specifies inclusion and exclusion criteria, outlines quality assessment, and develops a data extraction sheet, technical roadmap and Gantt chart, giving a clear path to answer the main research question.

The next stage will follow the PRISMA process: multi-database searches, two-stage screening by two reviewers, full-text quality checks, and structured data extraction covering clinical setting, integration models, outcomes, and ethics and privacy. Where results are comparable, meta-analysis or meta-synthesis will be done; where they are too different, an evidence map will be prepared. The study will also build a taxonomy that links AI methods, XR platforms, and integration modes with clinical outcomes, and assess both the maturity and the clinical effect of these technologies in practice. The final goal is to produce a clear, transparent, and reusable synthesis that shows the trends in AI+XR and provides the evidence needed for their safe and routine use in healthcare.

References

[1] M. J. Liberatore and W. P. Wagner. Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality*, 25(3):773–799, 2021.

[2] P. Milgram and F. Kishino. A taxonomy of mixed reality visual-displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, 1994.

[3] P. Aristeidis, K. Lefteris, K. Eleni, and K. Haridimos. Transforming healthcare: A comprehensive review of augmented and virtual reality interventions. *Sensors*, 25(12):3748–3765, 2025.

[4] B. Kyaw, N. Saxena, P. Posadzki, J. Vseteckova, C. Nikolaou, P. George, U. Divakar, I. Masiello, A. Kononowicz, N. Zary, and L. Tudor Car. 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, 2019.

[5] Davide Castelvecchi. Can we open the black box of ai? *Nature*, 538(7623):20, 2016.

[6] G. Lampropoulos. Intelligent virtual reality and augmented reality technologies: An overview. *Future Internet*, 17(2):58, 2025.

[7] A. D. Pelosi, N. Roth, T. Yehoshua, D. Itah, O. Braun Benyamin, and A. Dahan. Personalized rehabilitation approach for reaching movement using reinforcement learning. *Scientific Reports*, 14(1):17675, 2024.

[8] M. Sharma, C. Savage, M. Nair, I. Larsson, P. Svedberg, and J. M. Nygren. Artificial intelligence applications in health care practice: Scoping review. *Journal of Medical Internet Research*, 24(10):e40238, 2022.

[9] H. Ghaednia, M. S. Fourman, A. Lans, K. Detels, H. Dijkstra, S. Lloyd, A. Sweeney, J. H. F. Oosterhoff, and J. H. Schwab. Augmented and virtual reality in spine surgery, current applications and future potentials. *The Spine Journal*, 21(10):1617–1625, 2021.

[10] K. Selvan, M. Mina, H. Abdelmeguid, M. Gulsha, A. Vincent, and A. Sarhan. Virtual reality headsets for perimetry testing: A systematic review. *Eye*, 38(6):1041–1064, 2024.

[11] M. Mergen, M. Meyerheim, and N. Graf. Reviewing the current state of virtual reality integration in medical education – a scoping review protocol. *Systematic Reviews*, 12(1):97, 2023.

[12] E. A. Chance. The combined impact of ai and vr on interdisciplinary learning and patient safety in healthcare education: A narrative review. *BMC Medical Education*, 25(1):1039, 2025.

[13] A. Tabassum, I. Ghaznavi, A. Abd-Alrazaq, and J. Qadir. Exploring the application of ai and extended reality technologies in metaverse-driven mental health solutions: Scoping review. *Journal of Medical Internet Research*, 27(1):e72400, 2025.

[14] Z. Rudnicka, K. Proniewska, M. Perkins, and A. Pregowska. Health digital twins supported by artificial intelligence-based algorithms and extended reality in cardiology. *arXiv*, 2024. Preprint.

[15] Y. Wu, K. Hu, D. Z. Chen, and J. Wu. Ai-enhanced virtual reality in medicine: A comprehensive survey. *arXiv*, 2024. Preprint.

[16] Daniel T. Felsberg, Reza Pousti, Charles K. Howard, Sarah E. Ross, Laura D. Raisbeck, John T. McGuirt, and Christopher K. Rhea. The landscape of virtual reality use in mobility rehabilitation from 2010–2023: A scoping review. *Technologies*, 13(5):167, 2025.

[17] Nicholas Raison, Peter Harrison, Takashi Abe, Asim Aydin, Khurshid Ahmed, and Prokar Dasgupta. Procedural virtual reality simulation training for robotic surgery: A randomised controlled trial. *Surgical Endoscopy*, 35(12):6897–6902, 2021.

[18] Michael E. Ivan, David G. Eichberg, Leo Di, Ashish H. Shah, Ezequiel M. Luther, Vincent M. Lu, Ricardo J. Komotar, and Timur M. Urakov. Augmented reality head-mounted display–based incision planning in cranial neurosurgery: a prospective pilot study. *Neurosurgical Focus*, 51(2):E3, 2021.

[19] Christopher A. Molina, Christopher F. Dibble, Shawn L. Lo, Timothy Witham, and Daniel M. Sciubba. Augmented reality–mediated stereotactic navigation for execution of en bloc lumbar spondylectomy osteotomies. *Journal of Neurosurgery: Spine*, 2021.

[20] Hao Shi, Kai Zhang, Yong Li, Huan Shi, Xin Wen, and Tao Qian. The efficacy of augmented reality technology assisted by 3d-ct reconstruction in microvascular decompression for hemifacial spasm craniotomy. *BMC Surgery*, 25(1):382, 2025.

[21] Nancy J. Wei, Brian Dougherty, Amanda Myers, and Sherif M. Badawy. Using google glass in surgical settings: Systematic review. *JMIR mHealth and uHealth*, 6(3):e54, 2018.

[22] Hiroki Ase, Kei Honaga, Masayuki Tani, Taku Takakura, Fumiya Wada, Yuta Murakami, Ryo Isayama, Akira Tanuma, and Takahiro Fujiwara. Effects of home-based virtual reality upper extremity rehabilitation in persons with chronic stroke: A randomized controlled trial. *Journal of NeuroEngineering and Rehabilitation*, 22(1):20, 2025.

[23] Gyu Young Shim, Eun Hye Kim, Young Jun Baek, Won Kyung Chang, Byung Rim Kim, Jong Hoon Oh, Jung Il Lee, Jong Hyeok Hwang, and Ji-Young Lim. A randomized controlled trial of postoperative rehabilitation using digital healthcare system after rotator cuff repair. *npj Digital Medicine*, 6(1):95, 2023.

[24] Wei Deng, Dong Hu, Shuaipeng Xu, Xia Liu, Jing Zhao, Qian Chen, Jing Liu, Zhi Zhang, Wei Jiang, Liang Ma, Xia Hong, Shan Cheng, Bo Liu, and Xiang Li. The efficacy of virtual reality exposure therapy for ptsd symptoms: A systematic review and meta-analysis. *Journal of Affective Disorders*, 257:698–709, 2019.

[25] Daniel Freeman, Peter Haselton, Jason Freeman, Bernhard Spanlang, Sukhdeep Kishore, Emma Albery, Mark Denne, Paul Brown, Mel Slater, and Alastair Nickless. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, 5(8):625–632, 2018.

[26] Hunter G. Hoffman, Rosa A. Rodriguez, Mayra Gonzalez, Maria Bernardy, Roberto Peña, William Beck, David R. Patterson, and William J. Meyer. Immersive virtual reality as an adjunctive non-opioid analgesic for predominantly latin american children with large severe burn wounds during burn wound cleaning in the intensive care unit: A pilot study. *Frontiers in Human Neuroscience*, 13:262, 2019.

[27] Yan Wang, Jie Sun, Ke Yu, Xia Liu, Long Liu, Hui Miao, and Tao Li. Virtual reality exposure reduce acute postoperative pain in female patients undergoing laparoscopic gynecology surgery: A randomized control trial (rct) study. *Journal of Clinical Anesthesia*, 97:111525, 2024.

[28] Cristóbal Carvajal-Parodi, Patricio O. Rossel, Andrés Rodríguez-Alvarado, Francisco Guede-Rojas, and Jesús G. Ponce-González. Effects of virtual reality-based interventions on pain catastrophizing in people with chronic pain: A systematic review and meta-analysis. *Journal of Clinical Medicine*, 14(11):3782, 2025.

[29] G. Morone, L. De Angelis, A. Martino Cinnera, R. Carbonetti, A. Bisirri, I. Ciancarelli, M. Iosa, S. Negrini, C. Kiekens, and F. Negrini. Artificial intelligence in clinical medicine: A state-of-the-art overview of systematic reviews with methodological recommendations for improved reporting. *Frontiers in Digital Health*, 7, 2025.

[30] E. H. Shortliffe, S. G. Axline, B. G. Buchanan, T. C. Merigan, and S. N. Cohen. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Computers and Biomedical Research*, 6(6):544–560, 1973.

[31] H. B. Clark, J. Egger, and V. G. Duffy. Ai in healthcare and medicine: A systematic literature review and reappraisal. In V. G. Duffy, editor, *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pages 251–270. Springer Nature Switzerland, 2024.

[32] Bram van Ginneken, Arnaud A. A. Setio, Colin Jacobs, and Francesco Ciompi. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 286–289, 2015.

[33] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and D. R. Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, 2016.

[34] Scott M. McKinney, Marta Sieniek, Varun Godbole, Jon Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mark Chesus,

Greg S. Corrado, Ara Darzi, Mojtaba Etemadi, Francisco Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sarah Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, and Shravya Shetty. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[35] J. G. Kim, S. Y. Ha, Y.-R. Kang, H. Hong, D. Kim, M. Lee, L. Sunwoo, W.-S. Ryu, and J.-T. Kim. Automated detection of large vessel occlusion using deep learning: A pivotal multicenter study and reader performance study. *Journal of NeuroInterventional Surgery*, 2024.

[36] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, and D. Tse. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6):954–961, 2019.

[37] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin. Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3):457–469, 2017.

[38] X. A. Nguyen, D. Ljuhar, M. Pacilli, R. M. Nataraja, and S. Chauhan. Surgical skill levels: Classification and analysis using deep neural network model and motion signals. *Computer Methods and Programs in Biomedicine*, 177:1–8, 2019.

[39] S. Khalid, M. Goldenberg, T. Grantcharov, B. Taati, and F. Rudzicz. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Network Open*, 3(3):e201664, 2020.

[40] B. Debnath, M. O'Brien, M. Yamaguchi, and A. Behera. A review of computer vision-based approaches for physical rehabilitation and assessment. *Multimedia Systems*, 28(1):209–239, 2022.

[41] Pranav Rajpurkar, Eric Chen, Oishi Banerjee, et al. Ai in health and medicine. *Nature Medicine*, 28:31–38, 2022.

[42] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25:44–56, 2019.

[43] D. Reiners, M. R. Davahli, W. Karwowski, and C. Cruz-Neira. The combination of artificial intelligence and extended reality: A systematic review. *Frontiers in Virtual Reality*, 2:721933, 2021.

[44] S.-Y. Chiou, L.-S. Liu, C.-W. Lee, D.-H. Kim, M. A. Al-masni, H.-L. Liu, K.-C. Wei, J.-L. Yan, and P.-Y. Chen. Augmented reality surgical navigation system integrated with deep learning. *Bioengineering*, 10(5), 2023.

[45] A. Pierre, R. L. Cazzato, R. Nitin, de M. Pierre, R. Chloé, G. Julien, C. Y. Philippe, S. Jean-Paul, and G. Afshin. Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: A pilot randomised clinical trial. *European Spine Journal*, 29(7):1580–1589, 2020.

[46] J. Wang, Y. Qin, Q. Wu, D. Zeng, X. Gao, Q. Wang, Z. Li, Y. Ni, H. Li, P. Zhang, J. Guo, W. Ma, M. Maitusun, X. Jin, M. Chen, L. Zhu, Q. Lu, Q. Chen, Y. Wu, and H. Li. An adaptive ai-based virtual reality sports system for adolescents with excess body weight: A randomized controlled trial. *Nature Medicine*, 31(7):2255–2268, 2025.

[47] A. M. Fazlollahi, M. Bakhaidar, A. Alsayegh, R. Yilmaz, A. Winkler-Schwartz, N. Mirchi, I. Langleben, N. Ledwos, A. J. Sabbagh, K. Bajunaid, J. M. Harley, and R. F. Del Maestro. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: A randomized clinical trial. *JAMA Network Open*, 5(2):e2149008, 2022.

[48] B. Giglio, A. Albeloushi, A. Kh. Alhaj, M. Alhantoobi, R. Saeedi, V. Davidovic, A. Uthamacumaran, R. Yilmaz, J. Lapointe, N. Balasubramaniam, T. Tee, A. M. Fazlollahi, J. A. Correa, and R. F. Del Maestro. Artificial intelligence–augmented human instruction and surgical simulation performance: A randomized clinical trial. *JAMA Surgery*, 2025.

[49] C.-F. Tsai, C.-C. Chen, E. H.-K. Wu, C.-R. Chung, C.-Y. Huang, P.-Y. Tsai, and S.-C. Yeh. A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:2124–2132, 2021.

[50] S. Oh, Y.-S. Joung, T.-M. Chung, J. Lee, B. J. Seok, N. Kim, and H. M. Son. Diagnosis of adhd using virtual reality and artificial intelligence: An exploratory study of clinical applications. *Frontiers in Psychiatry*, 15:1383547, 2024.

[51] A. Hudon, M. Beaudoin, K. Phraxayavong, S. Potvin, and A. Dumais. Enhancing predictive power: Integrating a linear support vector classifier with logistic regression for patient outcome prognosis in virtual reality therapy for treatment-resistant schizophrenia. *Journal of Personalized Medicine*, 13(12):1660, 2023.

[52] S. Duan, L. Valmaggia, A. J. Lawrence, D. Fennema, J. Moll, and R. Zahn. Virtual reality-assessment of social interactions and prognosis in depression. *Journal of Affective Disorders*, 359:234–240, 2024.

[53] E. Lendaro, C. K. Van der Sluis, L. Hermansson, L. Bunketorp-Käll, H. Burger, E. Keesom, C. Widehammar, M. Munoz-Novoa, B. E. McGuire, P. O'Reilly, E. J. Earley, S. Iqbal, M. B. Kristoffersen, A. Stockselius, L. Gudmundson, W. Hill, M. Diers, K. L. Turner, T. Weiss, and M. Ortiz-Catalan. Extended reality used in the treatment of phantom limb pain: A multicenter, double-blind, randomized controlled trial. *PAIN*, 166(3):571, 2025.

Part 2

The Research Paper

Towards Intelligent Immersive Healthcare: A Systematic Review on the Integration of AI with AR/VR in Medical Applications

Author: Xiumiao Zhang

Abstract

Artificial intelligence (AI) and extended reality (XR) are increasingly being integrated in healthcare, moving from conceptual exploration toward empirical studies in clinical and clinically relevant settings. This shift is giving rise to new forms of intelligent, immersive interaction for clinical work, rehabilitation, mental health, and patient care. However, existing reviews have either examined AI and XR separately or focused on their integration within a single medical domain, leaving limited understanding of how these technologies are integrated across clinically relevant healthcare applications. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we conducted a comprehensive search of PubMed, Scopus, and Web of Science for peer-reviewed empirical studies published between January 2020 and December 2025 that integrated AI with XR in clinical or clinically relevant tasks involving human participants. From 339 deduplicated records, 41 studies met the inclusion criteria. We synthesised where these systems are used, how AI and XR are integrated, and how mature the current evidence is. We find that AI-XR integration is concentrated in surgery, mental health, and rehabilitation, and that integration patterns are shaped primarily by clinical task structures rather than by specific AI methods or XR devices. Although reported outcomes are largely positive, the field remains dominated by early-stage clinical piloting rather than large-scale validation or routine deployment. Overall, AI-XR healthcare systems are moving beyond prototypes, but their real-world impact depends on task-specific integration, clinically meaningful interaction, and stronger evidence of implementation in care settings.

Index Terms: artificial intelligence, extended reality, augmented reality, virtual reality, mixed reality, systematic review, clinical applications, technology readiness, PRISMA.

1 Introduction

The digital transformation of healthcare has become a strategic priority for health systems worldwide. The World Health Organisation's Global Strategy on Digital Health 2020–2025 positions digital health as a way to improve health outcomes, strengthen health systems, and support more accessible and person-centred care [1]. Artificial intelligence [2], [3], immersive technologies [4], [5], wearable sensors [6], medical imaging [7], and interactive data systems are reshaping how clinical information is generated and used.

Within this broader change, two technological directions have drawn particular attention: artificial intelligence (AI) and immersive technologies, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), collectively referred to in this review as extended reality (XR). Here, AI refers to computational methods for interpreting data, recognising patterns, making predictions, and supporting decisions [2], [3], while XR refers to interactive technologies that either immerse users in virtual environments or overlay digital information onto the physical world [8].

AI has become important in healthcare because it improves efficiency, reduces repetitive manual work, and supports more consistent interpretation of complex clinical data [2], [3], [9]. Methods such as machine learning, deep learning, computer vision, and natural language processing now support tasks including prediction, classification, segmentation, risk modelling, and decision support [3], [9]. Their relevance spans many clinical areas, from diagnosis and surgery to rehabilitation, mental health, oncology, wound care, and patient monitoring [3], [9]. For example, AI can detect lesions, segment anatomical structures, estimate patient risk, classify behavioural patterns, or assess movement quality [3], [10]. Yet an accurate or efficient AI output is not automatically useful in practice. It still needs to be presented in a form that clinicians or patients can act on within the clinical workflow [11], [12].

XR technologies provide new ways to present and interact with medical information. VR can create controlled and repeatable environments in which patients or users complete rehabilitation, exposure-based intervention, assessment, or therapy-related tasks [13], [14], [15]. AR and MR can overlay digital information onto the physical world, with applications in surgical guidance, procedural support, and anatomical visualisation [4], [16], [17]. The value of these technologies lies in the fact that many clinical tasks are not purely cognitive — they are spatial and embodied. Surgeons rely on anatomical and instrument-level cues during operation; rehabilitation patients benefit from feedback that responds to their movement; mental-health interventions often depend on environments in which behaviour can be observed under controlled conditions. XR therefore offers a task environment in which AI outputs can become visible and actionable.

Many clinical tasks require both reliable data interpretation and effective interaction with the clinical environment. For example, AI helps interpret clinical data, while XR places this information in a spatial, interactive setting. When the two are integrated, AI can make XR systems more adaptive and data-driven, while XR can make AI outputs easier to understand and apply to clinical tasks [18]. The integration of AI and XR is therefore more than the sum of two technologies; it may create new ways for medical information to be analysed and used in practice.

Clinically relevant applications are an important setting for examining this integration because technical feasibility does not necessarily mean clinical value. An AI-XR system may work well as a prototype, but it can only contribute to healthcare when it supports real or clinically relevant tasks and can be used by clinicians, patients, or caregivers. Focusing on clinical applications, therefore, helps assess whether AI and XR are not only technically connected but also meaningfully connected to medical practice.

Although the integration of AI and XR is promising, it is still unclear how the two are integrated in clinically relevant applications. Are AI and XR used side by side, or do they form distinct technical and clinical relationships? Do these integration patterns change across medical domains? Clinical applications differ in their goals, users, workflows, data sources, interaction demands, and levels of risk. These differences may shape the roles AI and XR play, and how they are integrated within clinical systems. It also remains uncertain whether reported effects of integrated AI-XR applications translate into clinically meaningful benefits, and how close current systems are to real-world use. Examining both reported effects and clinical implementation maturity is therefore necessary [11], [12], [19]. Since this is an emerging field, understanding current trends and challenges is equally important.

To address these gaps, this study conducts a systematic review, reported following PRISMA guidelines, to map current evidence on the integration of AI with XR in clinically relevant empirical studies [20]. The core research question is: Where and how is the integration of AI and XR having, or trending towards, real-world impact in clinically relevant medical applications?

To address this question, this study is structured around the following sub-questions:

- What are the medical application areas where AI and XR are used together, and who are the primary users?
- What types of AI methods and XR platforms or devices are commonly used in the existing studies?
- How are AI and XR integrated to support clinically relevant applications?
- How effective is the integration of AI and XR in real medical practice, and how mature is its clinical implementation?
- What are the current research trends in the integration of AI and XR technologies in clinically relevant medical applications, and what are the key limitations or gaps in existing studies?

The remainder of the paper is organised as follows. Section 2 introduces the AI, XR, and AI-XR background needed to frame the review. Section 3 describes the systematic review methodology. Section 4 maps the included evidence across applications, users, technologies, integration modes, outcomes, and readiness. Section 5 discusses what these patterns imply for clinical translation, limitations, and future research. Section 6 concludes the paper.

2 Background

2.1 Artificial Intelligence in Medicine

In medicine, AI is commonly used as a set of computational methods for interpreting clinical data and supporting decision-making, assessment, or feedback. These methods include prediction, classification, segmentation, risk modelling, decision support, and automated assessment [2], [3], [9]. They are most visible where clinical work depends on images, signals, movement data, or repeated judgement. AI methods have been used to detect lesions in medical images, segment anatomical structures, estimate patient risk [3], analyse surgical workflow and performance [21], and assess movement quality in rehabilitation [10].

Medical imaging is one of the clearest areas where these functions have developed. In radiology, pathology, ophthalmology, endoscopy, and wound assessment, machine learning and deep learning models identify patterns in images and support screening or diagnosis [2], [3]. Surgical and procedural settings provide another source of data — tool movement, video recordings, and operative steps — that can be used for performance analysis or planning support [21]. In rehabilitation, computer vision, pose estimation, and motion analysis make patient movement more measurable and less dependent on subjective observation [10]. AI should therefore be treated as a set of clinical functions rather than as a diagnostic tool alone, supporting monitoring, assessment, feedback, and planning across the clinical pathway [9].

The main value of AI here is not the replacement of clinical judgement. AI reduces repetitive work, improves consistency, and structures information that would otherwise be difficult to review at scale [2], [3]. A model may flag a risk, outline a structure, classify a behaviour, or estimate a clinical outcome, helping clinicians focus attention and make decisions more efficiently.

These strengths do not solve implementation problems. High model performance does not always translate into clinical usefulness. Many systems remain difficult to interpret, and clinicians may not know why a model produced a particular output. This matters in diagnosis, treatment planning, and patient-facing care, where trust and explanation are central [11], [12]. AI models may perform well in one dataset but less well in another hospital, device, or patient population, and bias, data quality, privacy, and validation remain major concerns [11], [12].

Another limitation is that AI outputs are often separated from the clinical task environment. A prediction, segmentation, or risk score may be accurate, but it still needs to be delivered at the right time and in a form that clinicians or patients can use [11]. This is where the connection with XR becomes important: if AI provides analysis, XR can provide the space in which that analysis becomes visible and usable.

2.2 XR Technologies in Medicine

In medicine, XR technologies are used to support tasks that require more than information on a conventional screen. Many medical applications involve spatial orientation, repeated practice, controlled exposure, or interaction within a task environment [4], [5], [13], [14]. VR provides a simulated environment that is safe and repeatable; AR overlays digital information onto the physical world; MR combines virtual and real elements in a spatially registered way [8]. These technologies are useful because many medical tasks are not only cognitive but also spatial and embodied.

In surgery, XR has been used for training, planning, navigation, and intraoperative support [4], [16], [17]. VR allows surgeons or trainees to practise procedures in a simulated setting before working with patients [5]. AR and MR are most relevant when anatomical models, imaging information, or navigation cues need to be positioned in relation to the patient or operative field [4], [16], [17]. The rationale is especially strong in surgery, where decisions depend on spatial relationships between anatomy, instruments, and planned actions.

In rehabilitation, XR can make repetitive exercises more engaging and measurable. Patients practise movements in a virtual or augmented environment, receive visual feedback, and complete tasks that are more motivating than conventional exercises [13]. Its relevance is clearest when long-term adherence and repeated practice shape outcomes. In mental health, VR has been used to create controlled exposure environments and therapeutic scenarios [14], [15]. These environments allow patient responses to be observed in a safe and repeatable way. XR has also been explored in adjacent care-support contexts, although the evidence base varies by application area.

The same technology, therefore, takes on different roles depending on the clinical task: precision and spatial guidance in surgery [4], [16]; feedback, motivation, and repeated practice in rehabilitation [13]; controlled exposure and behavioural observation in mental health [14], [15].

The same task-facing strengths also reveal the limits of current XR systems. Many systems still rely on fixed content, manual setup, or simple visual display, and do not adapt to patient performance, clinical data, or changes in user behaviour. In surgery, inaccurate registration, limited field of view, latency, and hardware discomfort can reduce usefulness [4], [16], [17]. In rehabilitation and mental health, engagement alone does not guarantee lasting clinical benefit [13], [14], [15]. Some applications work well in a small study but remain difficult to use in routine care [19].

Immersion is therefore only part of the clinical argument. XR often needs data interpretation, personalisation, feedback, and evaluation within real clinical workflows to become clinically useful — areas where AI can add value, but only if the two technologies are integrated in a meaningful way [18].

2.3 Related Reviews and Remaining Gaps

Existing reviews have provided useful evidence on AI in medicine and on XR in medicine, but they typically examine the two technologies separately. Reviews of AI in healthcare tend to focus on diagnosis, prediction, segmentation, clinical decision support, or implementation of AI systems [3], [9], [11]. These reviews explain the value and limits of AI, but rarely examine how AI outputs are presented or used within immersive clinical environments. Reviews of XR in medicine, in contrast, focus on surgical training, rehabilitation, mental health, or medical education [4], [5], [13], [14], [16]. They show the value of immersive technologies, but pay less attention to whether the systems are supported by AI-based analysis, prediction, or feedback.

A smaller group of reviews has begun to examine AI together with immersive technologies. Their contribution is useful but still partial. One broad AI–XR systematic review is not specific to medical applications [18], while domain-focused reviews examine areas such as medical education or mental health rather than the full cross-domain clinical evidence base [22], [23]. Others focus on a single immersive technology, such as VR, without covering AR or MR. Current reviews therefore do not yet provide a clear cross-domain picture of how AI and XR are integrated across clinically relevant medical applications.

Less well synthesised is the relationship between the two technologies once they appear in the same medical application. It is unclear whether AI and XR are simply used side by side or whether they form distinct technical and clinical relationships, and how such patterns vary across medical domains with different task structures. Existing reviews have also rarely examined reported effects together with evidence maturity, study design, sample size, and implementation context [19]. Without these connections, it is difficult to judge how close current AI–XR applications are to real-world clinical impact.

The remaining gap is therefore not whether AI and XR appear in the same studies, but how they are integrated within clinically relevant tasks. A useful synthesis needs to link application domain, user group, AI function, immersive technology, integration relationship, reported outcomes, and evidence maturity. In this review, integration is defined as a functional connection between AI and XR within the same clinical task, system, workflow, or data pathway, such that the two components serve a common clinical objective. This review addresses the identified gap by examining clinical and clinically relevant empirical studies of AI–XR in medical applications under this definition.

3 METHODOLOGY

3.1 Research design

This study was designed according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines. It systematically reviewed the integration of AI with extended reality (XR) technologies, including AR, VR, and MR, in clinically relevant medical applications. The review focused on studies in which AI and XR were functionally connected within the same clinical or clinically relevant task, system, workflow, or data pathway. The main review question was: where and how is the integration of AI and XR having, or trending towards, real-world impact in clinically relevant medical applications?

The review followed the PRISMA 2020 reporting framework for record identification, duplicate removal, screening, full-text eligibility assessment, exclusion reporting, and final study selection [20]. Given the heterogeneity of the included studies, meta-analysis was not conducted; evidence was synthesised using narrative synthesis supported by descriptive statistics and visual evidence mapping.

3.2 Search strategy

The literature search was conducted on 24 February 2026 and covered PubMed, Scopus, and Web of Science, spanning biomedical research, clinical medicine, digital health, computer science, engineering, and interdisciplinary technology research. The search strategy used four concept groups: immersive technology terms, AI terms, medical or healthcare terms, and clinically relevant empirical study terms. The publication date range was limited to 1 January 2020 to 31 December 2025 to focus on recent clinically relevant evidence in a rapidly developing field.

The immersive technology group included "Augmented Reality", "Virtual Reality", "Mixed Reality", and "Extended Reality". The AI group included "Artificial Intelligence", "Machine Learning", "Deep Learning", and "Computer Vision". The medical and healthcare group included healthcare, "health care", patient*, clinic*, and medic*. The clinical or empirical study group included "clinical trial", "randomized controlled trial", "randomised controlled trial", "prospective trial", "prospective study", "controlled clinical trial", "single-arm interventional trial", and "pilot clinical trial". Database-specific syntax was used to preserve the same conceptual logic while respecting different indexing systems.

In Scopus, the search was conducted in the TITLE-ABS-KEY field using the four concept groups, with limits for publication years 2020–2025, English language, and article-type records while excluding reviews. In Web of Science, the search used the topic field (TS) with the same four concept groups. In PubMed, the search combined Title/Abstract terms, MeSH terms, and publication type filters, including MeSH terms for Virtual Reality, Artificial Intelligence, Machine Learning, and clinical trial-related publication types. PubMed was also limited to publications from 1 January 2020 to 31 December 2025 and excluded review, systematic review, and meta-analysis publication types.

In simplified form, the Boolean structure was: (XR terms) AND (AI/ML/deep-learning/computer-vision terms) AND (healthcare/clinical/medical/patient terms) AND (clinical trial, prospective, interventional, controlled, single-arm, or pilot study terms). The trial-related and prospective/interventional terms were used to increase retrieval of clinically relevant empirical studies, not to define the review as limited to formal clinical trials. The complete database-specific strings are documented in the source notes and should be retained for auditability.

3.3 Eligibility criteria

Eligibility criteria were predefined and are summarised in Table 1. In brief, included studies had to be peer-reviewed, English-language empirical studies in a medical or health domain, involve human participants or human-participant data, and functionally integrate AI with XR within the same clinically relevant task, system, workflow, or data pathway. Exclusions were applied to records without eligible AI-XR integration, outside the medical or health domain, without clinically relevant empirical evaluation, or consisting of reviews, commentaries, protocols, design-only papers, non-peer-reviewed records, non-English papers, or animal, cadaver, phantom, or simulation-only studies without human participant data.

These criteria ensured that the review captured integrated AI–XR systems in clinically relevant medical contexts rather than AI-only, XR-only, or purely conceptual technology studies. The term "clinical or clinically relevant empirical study" was interpreted broadly enough to include formal clinical trials where present, but not to restrict the review to randomised or controlled clinical trials only.

3.4 Screening process

The PRISMA flow document was used as the authoritative source for screening numbers and exclusion categories. A total of 433 records were identified from databases: PubMed (n = 235), Scopus (n = 131), and Web of Science (n = 67). No additional records were identified from other sources, citation searching, or grey literature. Before screening, 94 records were removed as duplicates, all identified by Covidence, leaving 339 records for title and abstract screening.

Table 1: Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Integration of AI with XR	No AI usage, or AI usage outside the scope; no XR usage, or immersive technology outside the scope; AI and XR used independently, with no interaction/integration
Medical / health domain	Non-medical / health domain
Clinical or clinically relevant empirical study	Non-clinical study or no eligible clinically relevant empirical evaluation
Human participants, or human participant data	Animal, cadaver, phantom, simulation-only studies, or no human participant data
Empirical study	Literature review, commentary, meta-analysis, design paper without evaluation, protocol, or incomplete study
Written in English	Written in other languages
Peer-reviewed	Non-peer-reviewed

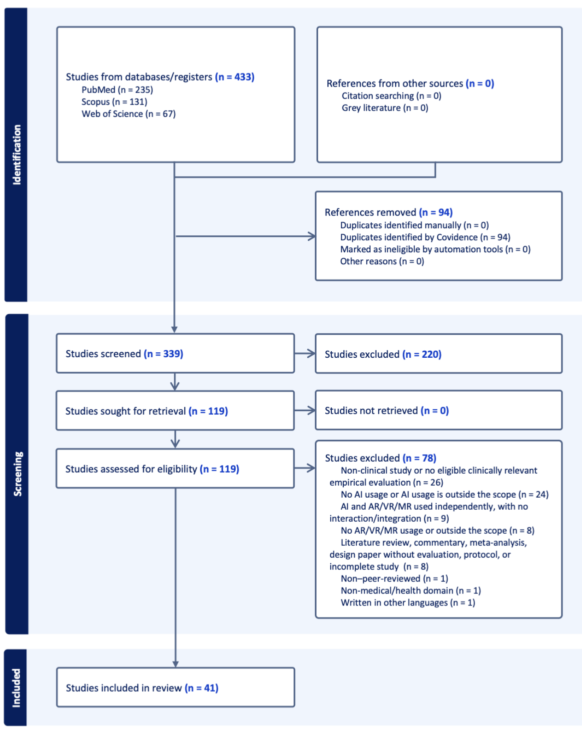

Figure 1: PRISMA flow diagram for study selection.

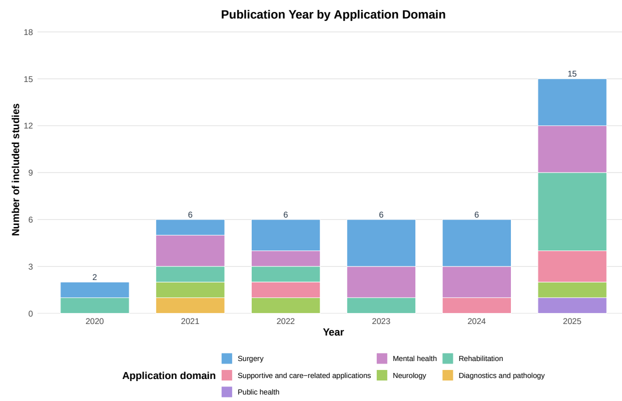

Figure 2: Publication year by application domain. The 2025 increase spans several domains, while surgery, mental health, and rehabilitation remain the most evaluated domains across the period.

At title and abstract screening, 220 records were excluded. The remaining 119 studies were sought for retrieval and assessed for eligibility at full text. No studies were recorded as not retrieved. During full-text assessment, 78 studies were excluded for the following reasons: non-peer-reviewed (n = 1); non-clinical study or no eligible clinically relevant empirical evaluation (n = 26); non-medical / health domain (n = 1); written in other languages (n = 1); no XR usage or outside the scope (n = 8); no AI usage or AI usage outside the scope (n = 24); AI and XR used independently, with no interaction/integration (n = 9); and literature review, commentary, meta-analysis, design paper without evaluation, protocol, or incomplete study (n = 8). Forty-one studies were included in the final review.

Screening was conducted in Covidence against the predefined eligibility criteria. Title/abstract screening and full-text assessment were performed independently by two reviewers. Disagreements were resolved by a third reviewer. Covidence was used as the complete screening management tool, including duplicate identification and screening workflow management.

4 Results

The Results follow the review questions, moving from the overall evidence base (Section 4.1) to application domains and target users (Section 4.2), the technical characteristics of the systems (Section 4.3), and the patterns of AI-XR integration that are the central focus of this review (Section 4.4). Section 4.5 reports outcome direction and technology readiness, interpreted cautiously, and Section 4.6 summarises the patterns that inform the Discussion.

4.1 Overview of the Evidence Base

The search and screening process produced 41 included clinical and clinically relevant empirical studies. Publication activity concentrated toward the end of the review period: two studies appeared in 2020 (4.9%), six studies were published in each year from 2021 to 2024 (14.6% each), and 15 studies appeared in 2025 (36.6%).

Figure 2 combines publication year and application domain, showing that the 2025 increase was not driven by a single area but distributed across surgery, rehabilitation, mental health, supportive care, neurology, and public health. Surgery was consistently represented throughout and contributed substantially to the 2025 rise; rehabilitation showed the most visible recent expansion; mental health appeared across several years without a steady year-on-year increase; and supportive-care applications emerged mainly in the later years. Public health remained an isolated example.

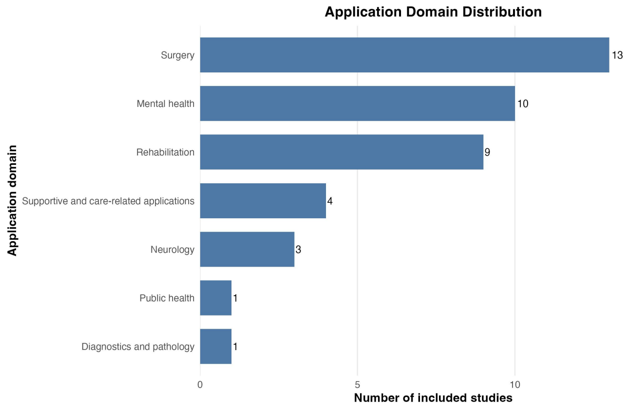

Figure 3: Application domain distribution. Surgery, mental health, and rehabilitation were the most evaluated domains; public health appears as an isolated example.

4.2 Application Domains and Target Users

As shown in Figure 3, the most frequently evaluated domains were surgery (n = 13, 31.7%), mental health (n = 10, 24.4%), and rehabilitation (n = 9, 22.0%). Supportive and care-related applications accounted for four studies (9.8%), neurology for three (7.3%), and diagnostics/pathology and public health for one study each (2.4% each). The field is therefore clinically oriented but unevenly distributed across healthcare.

The surgical cluster was shaped by spatial tasks such as 3D reconstruction, registration, planning, navigation, and intraoperative overlay. These appeared in automatic registration for laparoscopic liver surgery [31], VR/AI planning of lung segmentectomies [44], AR navigation for lumbar discectomy [39], and AI-3D-AR guidance in robotic prostatectomy [59]. Surgery is a natural fit for AI-XR because AI can segment anatomy or register models while XR makes that spatial information visible during planning and guidance.

The mental-health studies centred on immersive behavioural tasks paired with AI-based prediction, classification, or conversational support. VR-based stabilization for posttraumatic stress symptoms was combined with machine-learning prediction of symptom change [24], while social anxiety, spider phobia, and treatment-resistant schizophrenia were addressed through VR exposure, serious games, or avatar therapy, with AI used for prediction, adaptive exposure, or analysis of therapy transcripts [26], [28], [30], [46], [49], [53]. This domain used immersive environments not as visualisation tools but as structured contexts for eliciting behaviour, speech, physiology, or interaction data.

Rehabilitation formed a third cluster, linking AI-XR systems to performance monitoring, motor control, and adaptive feedback. Examples included AI-based assessment within VR hand-exoskeleton training for post-stroke rehabilitation [35], machine-learning-supported phantom motor execution with AR/VR feedback [40], and myoelectric VR serious gaming for chronic stroke [64]. Rehabilitation suits closed feedback loops because movement, muscle activity, and task performance can be captured and used to adjust therapy in real time.

These domains also raise a user-level question: are such systems built for clinicians, patients, or both? Target-user coding showed a strongly patient- and doctor-centred evidence base. Patients were the intended users in 25 studies (61.0%), doctors in 13 (31.7%), and mixed users in three (7.3%). Doctor-facing studies concentrated in surgery and diagnostic-support workflows, while patient-facing studies dominated mental health, rehabilitation, supportive care, and public health.

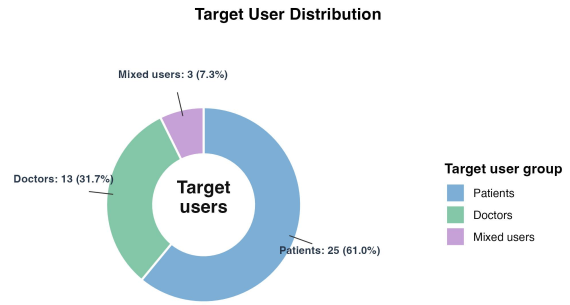

Figure 4: Target user distribution, showing a patient- and doctor-centred evidence base. This reflects the intended users, not the unit of analysis used for evaluation.

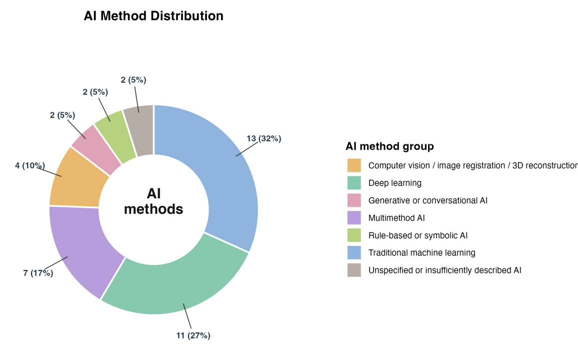

Figure 5: Proportional distribution of AI method families. These are technical approaches, not indicators of effectiveness.

4.3 Technical Characteristics of AI-XR Systems

The technical profile of the included systems was analysed at two levels. The AI method group describes the technical method family, such as traditional machine learning, deep learning, multimethod AI, or generative/conversational AI. The AI function group describes the role AI played within the AI-XR system, such as planning support, classification, prediction, coaching, or adaptive control. These two layers address different questions: the method group describes how AI was implemented technically, while the function group explains what AI contributed to the clinically relevant workflow.

Traditional machine learning and deep learning were the largest families of technical methods. Traditional machine learning appeared in 13 studies (31.7%), including predicting the response after VR-based stabilisation for posttraumatic stress symptoms [24] and classifying sarcopenia from Kinect-based MR exercise data [29]. Deep learning appeared in 11 studies (26.8%), with examples such as AI-empowered Ki67 pathology assessment and speech decoding with avatar control [45], [51]. Generative/conversational AI and rule-based/symbolic AI were smaller but relevant categories, appearing in AR-based mental-health support, MR exercise consultation, adaptive exposure, and serious-game therapy [28], [32], [38], [46].

AI function was more directly linked to the clinical role of each system. In surgery, AI most often supported navigation, planning, registration, or overlay, as seen in AR navigation for lumbar discectomy, AR-guided robotic prostatectomy, and multimodal AR planning for extracerebral tumours [39], [58], [59]. In rehabilitation, AI was more often used for adaptive feedback, motor control, or performance assessment, including hand-exoskeleton training, phantom motor execution, and myoelectric VR serious gaming [35],

Table 2: AI method families, system functions, and constituent studies.

AI methods	AI functions	Studies
Traditional machine learning	Classification / diagnosis / screening	[26], [29], [43], [48]
	Prediction / prognosis	[24], [30], [53]
	Adaptive personalization / closed-loop control	[35], [40], [61], [64]
	Performance or functional assessment	[49], [57]
Deep learning	Classification / diagnosis / screening	[45], [55]
	Segmentation / reconstruction / registration	[31]
	Navigation / planning support	[39], [44], [50], [52], [56], [58], [59]
	Adaptive personalization / closed-loop control	[51]
Computer vision / image registration / 3D reconstruction	Segmentation / reconstruction / registration	[27]
	Navigation / planning support	[34], [36], [41]
Rule-based or symbolic AI	Adaptive personalization / closed-loop control	[28]
	Coaching / tutoring / conversational support	[46]
Generative or conversational AI	Coaching / tutoring / conversational support	[32], [38]
Multimethod AI	Classification / diagnosis / screening	[25]
	Navigation / planning support	[54]
	Adaptive personalization / closed-loop control	[37], [63]
	Coaching / tutoring / conversational support	[62]
	Performance or functional assessment	[42]
	Generative content or guidance	[60]
Unspecified or insufficiently described AI	Adaptive personalization / closed-loop control	[33], [47]

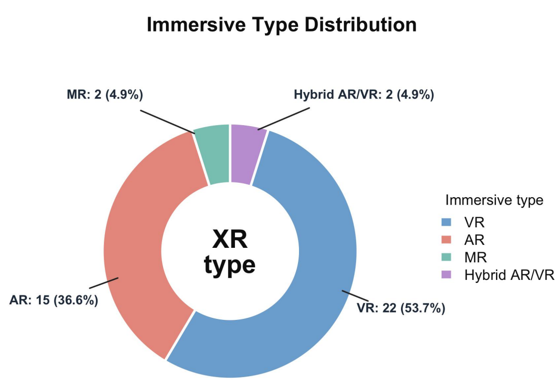

Figure 6: Immersive type distribution. VR and AR dominate the included studies, while MR and hybrid AR/VR are sparse.

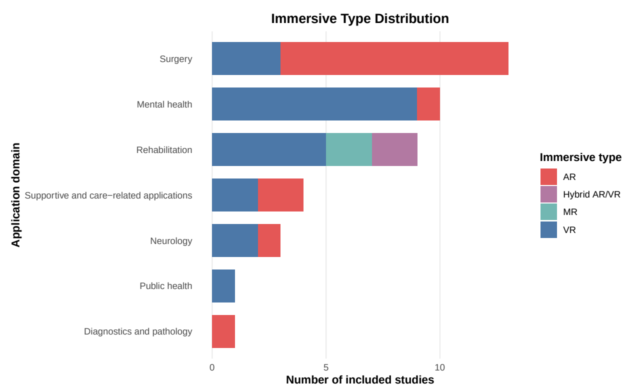

Figure 7: Immersive type by application domain. VR predominates in immersive mental-health and rehabilitation tasks, whereas AR concentrates in surgery and procedural support.

[40], [64]. In mental health, AI commonly supported prediction, classification, conversational/coaching support, or behavioural data analysis around VR exposure, stabilization, and avatar therapy [24], [26], [30], [49], [53]. In supportive care, AI-XR systems were used for preparation, monitoring, guidance, or care interaction, including paediatric endoscopy preparation and colostomy-care monitoring [60], [63].

The XR layer was dominated by VR (22 studies, 53.7%) and AR (15 studies, 36.6%), with MR and hybrid AR/VR in two each (4.9%; Figure 6). Examined by domain (Figure 7), a task-related pattern emerged. VR predominated where the task required an immersive environment, such as exposure therapy, social or avatar interaction, rehabilitation training, and behavioural intervention, whereas AR concentrated where the task involved spatial overlay, registration, or guidance, especially surgery and procedural support. MR and hybrid systems were sparse and appeared mainly in rehabilitation contexts. XR modality thus tracked the clinical task: VR to create an immersive environment, AR to place digital information into physical or procedural space.

The type of XR hardware used in the included studies was related to both XR modality and clinical setting. VR studies were relatively consistent in their device choices. Most used commercial head-mounted displays, especially the HTC Vive/Vive Pro Eye and Oculus/Meta Quest or Rift series [24], [25], [27], [37], [43], [55], [62]. A smaller number used older headsets, smartphone-based VR, screen-based VR, serious-game simulators, or custom rehabilitation setups [30], [33], [34], [42], [46], [47], [51], [64].

AR studies showed greater hardware diversity and were more often tied to clinical or procedural use. These included surgical navigation and guidance systems [31], [39], [59], robot-assisted AR platforms with fluorescence-imaging overlay [36], [54], mobile- or tablet-based AR systems [48], [50], [56], [57], [58], and projector- or camera-based AR systems [41], [63]. Only one AR study used wearable AR headsets, including Microsoft HoloLens and Apple Vision Pro [32]. MR and hybrid AR/VR appeared in only a few studies and were mainly represented by headset-, sensor-, or screen-based rehabilitation systems [29], [38], [40], [61].

4.4 Patterns of AI-XR Integration Across Clinical Domains

After completing the screening process and analysing the included studies in detail, this review grouped AI-XR integration into three modes based on the direction of information flow, the timing of

Table 3: XR hardware by immersive type and device category.

XR-type	Device category	Study
VR	Head-mounted display (HMD)	[24]-[28], [30], [33], [35], [37], [43], [44], [53], [55], [60], [62]
	Screen-based VR	[34], [46], [51], [64]
	VR rehabilitation platform	[42], [47]
	Not clearly reported	[49]
AR	Mobile or tablet-based AR	[48], [50], [56]-[58]
	AR surgical navigation/guidance system	[31], [39], [59]
	Projector/camera-based AR system	[41], [63]
	Robot-assisted AR platform	[36], [52], [54]
	Head-mounted display (HMD)	[32]
	AR microscope platform	[45]
MR	Head-mounted display (HMD)	[38]
	Sensor-based mixed-reality system	[29]
Hybrid AR/VR	Hybrid AR/VR rehabilitation system	[40]
	Screen-based AR/VR	[61]

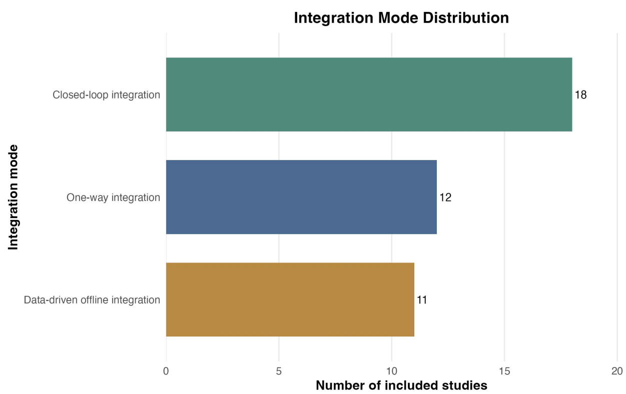

Figure 8: Integration mode distribution. Closed-loop is the largest category, but one-way and data-driven offline integration remain substantial.

AI analysis, and the way AI and XR worked together within the clinical task. These modes were one-way integration, data-driven offline integration, and closed-loop integration.

One-way integration refers to systems in which AI outputs flow into the AR/VR display, but user actions within XR do not feed back to change the AI component. In this mode, AI may generate outputs such as segmentation, reconstruction, registration, prediction, planning, or guidance, and AR/VR presents or uses these outputs through an immersive or spatial interface.

Data-driven offline integration refers to systems in which VR/AR collects or structures behaviour, image, speech, physiological, clinical, or performance data, and AI then performs diagnosis, classification, prediction, assessment, or modelling after or outside the immersive session. The AI results are not fed back into the XR environment in real time.

Closed-loop integration refers to systems in which VR/AR provides an adaptive training or interaction environment, users or patients perform tasks, AI analyses their behaviour, performance, physiological signals, or other data, and the VR/AR scene or feedback then updates accordingly during use.

Closed-loop integration was the largest category (n = 18, 43.9%), followed by one-way integration (n = 12, 29.3%) and data-driven offline integration (n = 11, 26.8%). This distribution shows that many included systems moved beyond simple AI-to-XR display, but closed-loop integration was not evenly distributed across domains.

The overall distribution does not show how the modes relate to system components. Figure 9 places immersive type, AI method family, and integration mode along a single pathway to examine how they co-occur.

VR connected to several AI method families, including traditional machine learning, deep learning, and multimethod AI, while AR linked more often to vision-oriented methods such as deep learning and computer vision/registration. The same AI method family, however, appeared across different integration modes, so integration mode was not determined by the AI method but by how each system connected AI and XR within its clinical task. The organising axis is therefore the structure of the clinical task, not the underlying technology stack.

Because clinical tasks differ across domains, the way AI and XR combined also varied by area, as summarised in the domain-by-mode heatmap (Figure 10) and the examples below.

Surgery was the clearest case of one-way integration: AI generated or registered spatial information that AR or VR then presented for planning or navigation, as in laparoscopic liver registration [31], lung segmentectomy planning [44], lumbar discectomy navigation [39], and multimodal AR planning for extracerebral tumours [58]. Some surgical systems were coded closed-loop when registration or overlay updated dynamically, such as ICG-driven AR robotic partial nephrectomy [36] and AR/AI vertebroplasty navigation [41].

Rehabilitation showed a different architecture, strongly associated with closed-loop systems in which AI contributed to assessment, motor-intent decoding, assistance-as-needed, or adaptive feedback. Examples include VR hand-exoskeleton rehabilitation with real-time assessment and adaptive task difficulty [35], home-based phantom motor execution with machine-learning decoding [40], and myoelectric VR serious gaming for chronic stroke [64]. Mental health, by contrast, often used XR to generate behavioural, speech, or physiological data analysed offline, as in posttraumatic-stress response prediction [24], anxiety prediction from VR sessions [26], and avatar-therapy transcript analysis [49], [53].

These cross-domain differences are central to the review question. The single label AI-XR integration covered quite different architectures depending on the clinical task. Integration mode, therefore, describes how the technologies are integrated, not how mature the evidence is. Whether these architectures are supported by mature evidence, instead of by positive early reports alone, is examined next.

4.5 Reported Outcomes and Technology Readiness

Reported result direction was predominantly positive: 39 studies (95.1%) were coded positive, one mixed (2.4%), and one limited or

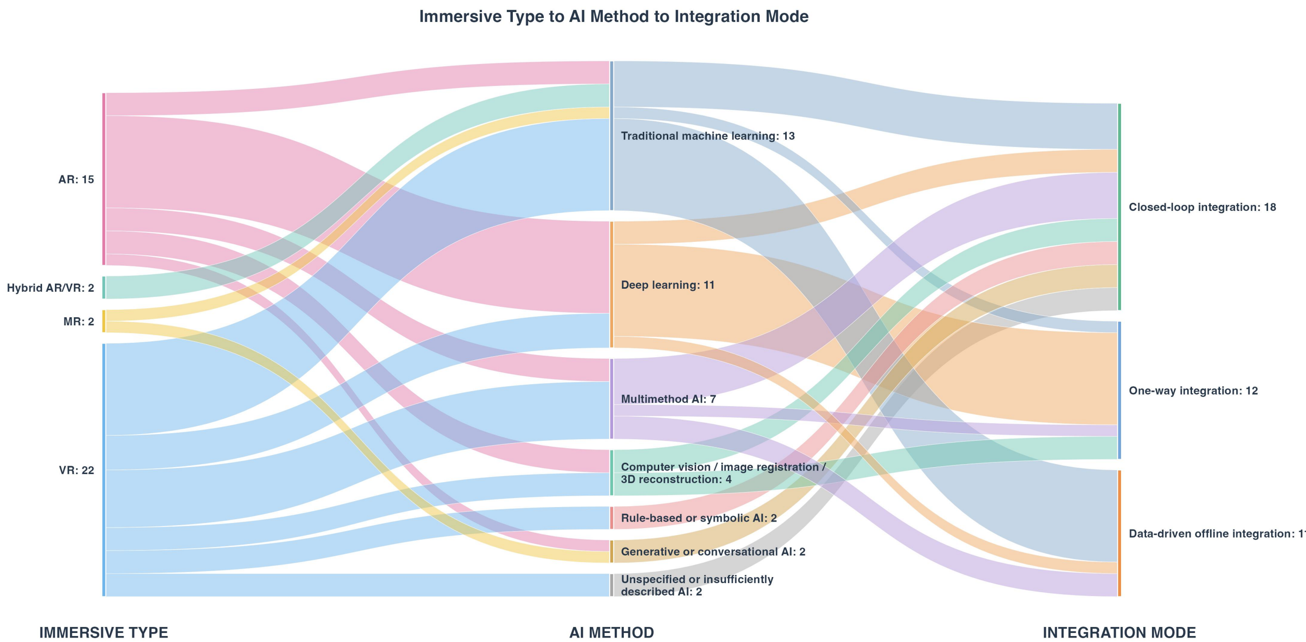

Figure 9: Alluvial mapping of immersive type, AI method family, and integration mode; flow width represents included-study counts.

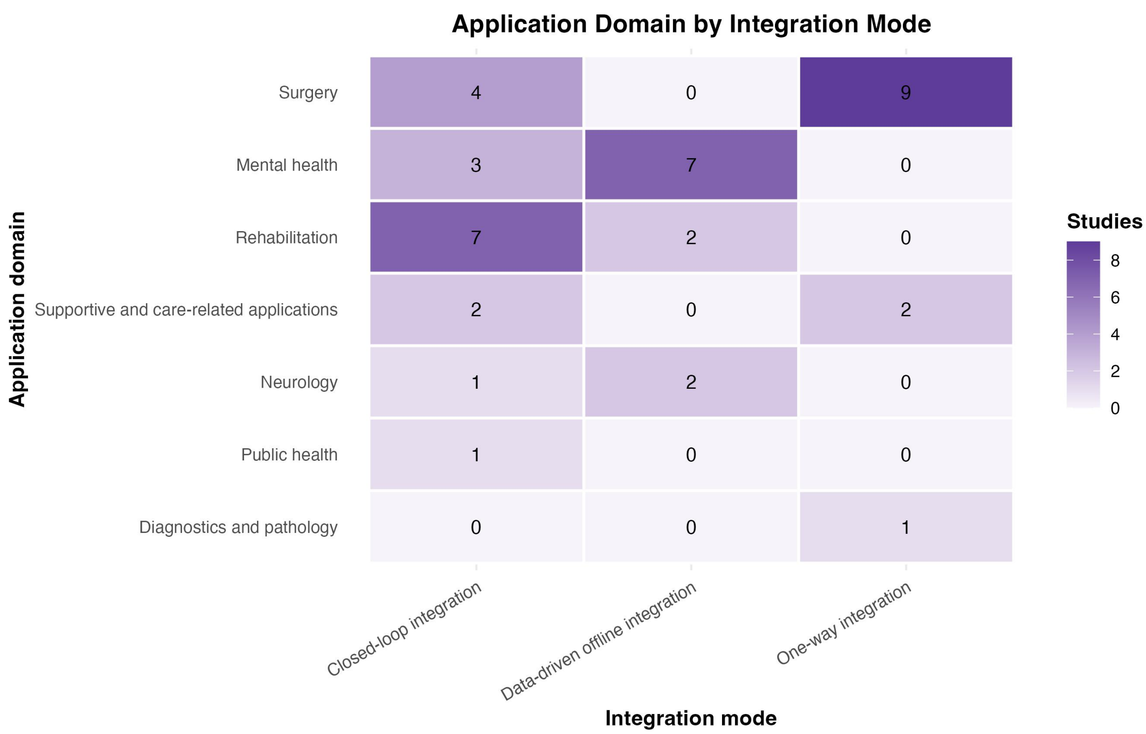

Figure 10: Application domain by integration mode. Surgery clusters around one-way spatial guidance, rehabilitation around closed-loop adaptation, and mental health around data-driven offline modelling with some closed-loop intervention.

inconclusive (2.4%). The positive findings spanned technical performance, feasibility or usability, symptom or behavioural change, clinical-workflow measures, and classification or prediction performance.

This near-uniform positivity must be read cautiously, because the code reflects only the direction of reported findings, not their strength, and several positive studies reported mixed detail beneath the overall label. The AR lumbar-discectomy navigation study reported fewer puncture attempts and fluoroscopy shots but no significant difference in operation time or several clinical scores [39]. The AI-3D-AR prostatectomy trial reported lower positive-margin rates and other favourable outcomes, but within one surgical context and design [59]. The extended-reality phantom-limb-pain trial showed improvement in both treatment arms, but phantom motor execution was not superior to phantom motor imagery, hence its mixed coding [61]. The spider-phobia study showed strong group-level benefit from one-session VR exposure but limited, non-significant longer-term prediction, explaining the limited/inconclusive coding of its AI component [30].

Because most studies reported positive findings, the more discriminating question is how mature the underlying technology and evidence were. Each system was interpreted against sector-based technology-readiness sources for health, care, digital, and pathway innovations [65], [66], with digital-health evidence standards used to frame evidence maturity separately [19]. In these sources, technology readiness levels (TRLs) describe how developed an innovation is, from validation in relevant settings through small-scale piloting, larger or operational evaluation, and routine adoption; for digital and pathway innovations, TRLs are flexible and should be interpreted in context [65], [66]. Because AI-XR systems combine software and device components, TRL was treated as an indicator of development and evaluation maturity, not of clinical effectiveness or routine deployment.

The extracted TRL values concentrated in the middle of this pathway (Table 4). TRL 6 was by far the most common level (27 studies, 65.9%). TRL 5 applied to three studies (7.3%) and TRL 7 to eleven studies (26.8%). No study reached the higher levels of testing at scale with cost-effectiveness evaluation or routine adoption into care pathways. Higher TRL did not automatically indicate operational maturity: comparatively strong evaluations such as the adaptive VR sports trial in adolescents with excess body weight [62], colostomy-care monitoring with AR and deep learning [63], and AI-3D-AR robotic prostatectomy [59] provided stronger clinical-evaluation evidence than small pilots without demonstrating routine deployment. Cross-tabulating integration mode against TRL (Table 5) shows that this maturity was not evenly shared across architectures. Data-driven offline integration did not reach TRL 7, remaining at TRL 5 and 6 only, while one-way integration (5 of 12 studies) and closed-loop integration (6 of 18) both progressed to larger-cohort clinical testing. Evaluation maturity thus mirrored the integration architecture: offline-modelling systems, concentrated in mental health, have so far been evaluated mainly at the feasibility stage. Surgical one-way guidance and rehabilitation closed-loop systems more often advanced to larger clinical cohorts, although none progressed beyond TRL 7.

These readiness differences can be read alongside the timing, application domain, and scale of each study in a combined evidence map (Figure 11), which plots every included study by publication year and application domain, with bubble colour indicating integration mode and bubble size indicating the comparable human, clinical-user, or patient-derived sample available for evaluation.

The map includes 39 studies with comparable sample-size values and excludes two with non-comparable units. Larger evaluations tended to use controlled or randomised designs, including the adolescent VR sports trial [62], colostomy monitoring [63], extracerebral tumour planning [58], and AI-3D-AR robotic prostatectomy [59]. Several technically sophisticated systems, by contrast, were tested only in small feasibility or case-based samples, such as the speech neuroprosthesis with avatar control [51], lung segmentectomy planning [44], and home-based phantom motor execution [40]. Larger samples contributed to evidence strength but were not sufficient on their own to establish maturity or general clinical effectiveness. Overall, the predominantly positive results rest on an evidence base concentrated at real-setting piloting, unevenly scaled, and not yet supported by routine-deployment evidence.

5 Discussion

5.1 Principal findings

This review examined where and how the integration of AI with AR and VR is approaching real-world impact in clinically relevant medical applications. Three findings stand out. First, AI-XR integration is no longer speculative: it is being tested with patients, clinicians, and clinical data across a widening set of medical tasks, with study volume rising sharply in 2025. Second, integration patterns are shaped more by clinical tasks than by AI methods or display hardware, meaning that the same broad technology label covers different forms of technical and clinical connection. Third, research activity should not be mistaken for clinical maturity: reported outcomes were almost uniformly positive, but the evidence remained concentrated at small-scale piloting, heterogeneity prevented meta-analysis, and no study reached testing at scale or routine adoption. The following sections discuss how clinical tasks shape integration architecture, why positive findings should be interpreted cautiously, what ethical and practical challenges arise from combining AI with immersive systems, and what these patterns imply for future AI-XR healthcare research.

5.2 Integration is organised by the clinical task, not by the technology

A central analytical result of this review is the separation of one-way, data-driven offline and closed-loop integration, and the finding that these modes followed the clinical task. Surgery favoured one-way integration. AI reconstructed, segmented or registered anatomy, and AR or VR then presented it for planning and navigation [31], [44], [58]. A few surgical systems closed the loop, but only where the overlay updated during the procedure itself [36], [54]. Rehabilitation was dominated by closed-loop designs, because movement, muscle activity and task performance can be sensed continuously and used to adapt the task as the patient works [35], [40], [64]. Mental-health studies more often used the immersive environment to draw out behavioural, physiological or linguistic data, which AI then analysed after the session [24], [30], [49]. The alluvial analysis made the point clearer. The same AI method family appeared across several integration modes, so the architecture of an AI-XR system reflects the logic of the clinical problem more than the algorithm chosen to address it. This pattern is established only in the domains with sufficient studies: surgery, rehabilitation, and mental health. Diagnostics and public health each contributed a single study, too few to support any claim about their integration mode.

Surgery's persistence in the one-way mode is worth noting. Intra-operative decisions carry immediate risk, and a system that allowed an AI model to alter guidance autonomously would face safety and regulatory demands that few current tools can meet. The cautious pattern observed here, in which AI prepares or registers information that a surgeon then interprets, may be appropriate rather than a sign of immaturity. The case for moving surgery towards closed-loop control needs to be made task by task.

This changes how the field should be read. The umbrella term AI-XR covers systems that do very different things, from guiding a surgeon's hands, to scoring a stroke patient's movement, to

Table 4: Distribution of included studies across technology readiness levels (TRLs).

TRL	Meaning in a health / digital health context	Number of studies	Percentage
TRL 5	Validated in a relevant environment using human participant-related data, but not directly evaluated with human participants.	3	7.3%
TRL 6	A near-final prototype is tested in a small cohort of human participants for performance, usability, safety, characterisation, and reproducibility.	27	65.9%
TRL 7	The system is tested in a larger cohort of human participants, often in a realistic clinical or operational setting. Usability, safety, and performance are closely monitored.	11	26.8%

Table 5: AI-XR integration mode by technology readiness level.

AI-XR integration mode	TRL 5	TRL 6	TRL 7	Total
Closed-loop integration	1	11	6	18
One-way integration	1	6	5	12
Data-driven offline integration	1	10	0	11
Total	3	27	11	41

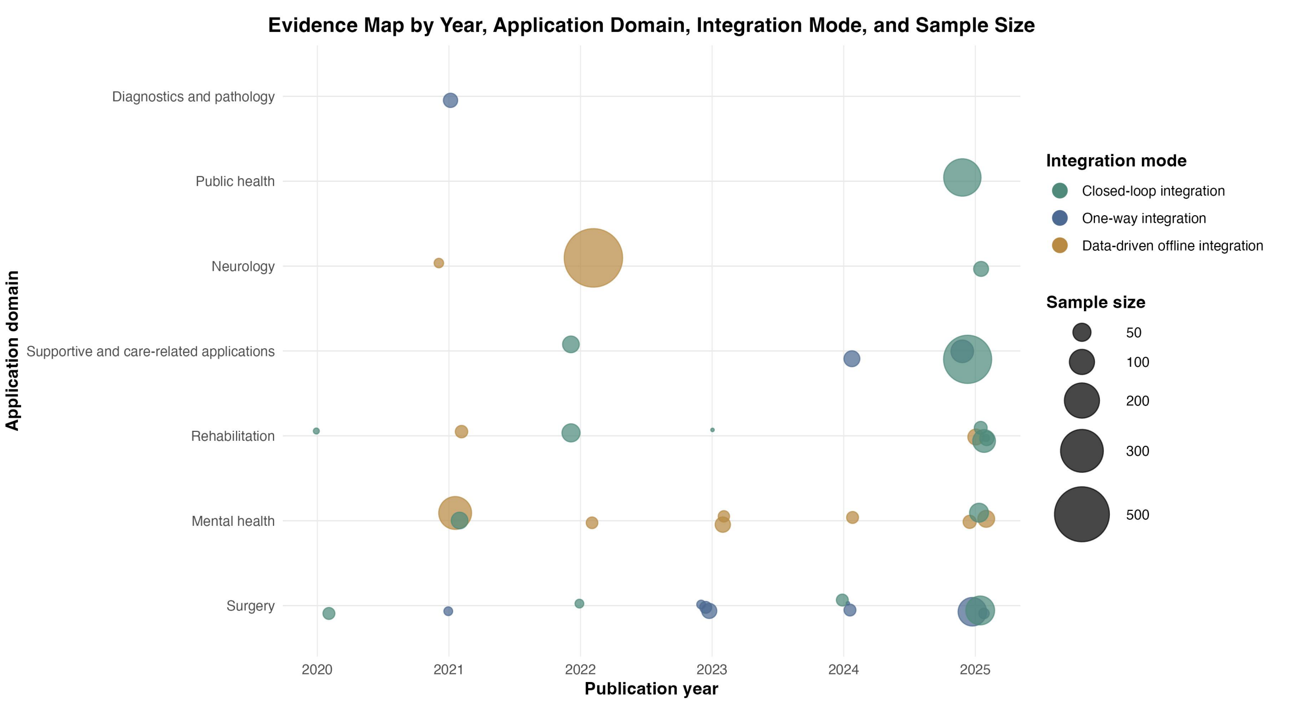

Figure 11: Evidence map by year, application domain, integration mode (bubble colour), and sample size (bubble size, comparable human or patient-derived counts where extractable).

modelling a phobia from session data. Treating them as one category obscures more than it clarifies. Earlier reviews that examined AI and immersive technology separately, or studied their combination outside medicine [18] or within a single domain such as mental health [22] or healthcare education [23], could not surface this structure. Organising the evidence by integration mode and clinical task gives a more faithful basis for comparison and for judging where deeper, closed-loop integration is actually warranted rather than only possible.

5.3 From positive signals to clinical impact

Set against the main question, the evidence is clinically oriented but still some distance from demonstrated impact. The systems were overwhelmingly patient-facing and clinician-facing, and 39 of 41 studies reported positive results. That near-uniformity is a reason for caution rather than confidence. It is consistent with an early-stage literature dominated by feasibility work, and with the familiar tendency to publish favourable findings, and it sits awkwardly beside the maturity data. Several studies coded as positive carried more guarded detail beneath the headline, including non-significant differences on secondary clinical outcomes [39], no advantage over a simpler comparator [61], and weak longer-term prediction [30]. That detail disappears if the direction of a result is taken as proof of benefit.

As in clinical AI more broadly [11], [12], strong technical performance and promising pilots do not show that a system improves care across real workflows, varied patients and sustained use. For AI-XR, this gap is compounded by hardware dependence, cost, discomfort and the difficulty of fitting real-time systems into clinical routines. Current evidence therefore reports promise and feasibility far more often than demonstrated effectiveness.

The evidence map pointed to a related issue about scale. The larger evaluations clustered in a handful of controlled or randomised studies [59], [62], [63]. Several of the most technically ambitious systems, including the speech neuroprosthesis with avatar control [51] and home-based phantom motor execution [40], were reported in only a few cases. Small samples are reasonable at the proof-of-concept stage, but they limit what any single study can claim.

5.4 Ethical, privacy, safety and accountability considerations

Integrating AI with immersive technology raises concerns that are sharper than those of either technology used on its own. Immersive systems collect rich and often intimate data, including gaze, movement, voice, physiological responses and behaviour during exposure or therapy tasks, and AI converts these into clinical inferences. The resulting record is more revealing than a conventional one, yet questions of consent, storage and secondary use received little attention in the included studies. Closed-loop systems concentrate the difficulty. When an AI component changes what a patient sees, or alters a surgical overlay in real time, responsibility for an error becomes harder to divide among clinician, developer and model. Interpretability and trust, already central to clinical AI [11], grow more complicated when outputs are embedded in an immersive scene that can feel seamless and therefore authoritative. Access is a further concern. Reliance on headsets and specialised hardware could widen disparities, although the same technology can also narrow them, as in the use of VR to bring surgical expertise to a low-income setting [27]. Systems that are at once software and medical device sit awkwardly within existing regulatory categories, and the absence of any near-deployment study suggests that safety monitoring and accountability arrangements for such hybrids remain underdeveloped.

Two issues are specific to the immersive setting. Immersive experiences can themselves cause harm, from cybersickness to the distress of confronting feared stimuli during exposure, and the included studies rarely reported how such effects were monitored or managed. Many of the applications also involved vulnerable groups, including children and patients with dementia, stroke or serious mental illness, for whom comprehension, valid consent and the persuasive force of immersive content are especially pressing. Bias in the AI component compounds these risks, since a model trained on narrow data may perform unevenly across the very populations these systems are intended to serve.

5.5 Research trends and gaps in the current evidence

Two related trends shape the evidence base. The first is breadth. After a stable volume from 2021 to 2024, study numbers rose sharply in 2025, and the work moved well beyond its imaging-heavy surgical origins into behavioural therapy, rehabilitation and supportive care [27], [62], [63]. The second is depth. Closed-loop designs, in which AI adapts the immersive experience in real time, were the most common mode, and recent work has increasingly combined several AI components or introduced generative and conversational elements instead of relying on a single classifier. These developments suggest growing depth in AI-XR integration, in the sense that the connection between AI and XR is becoming more dynamic, multi-component, and embedded within clinical task environments.

Against this momentum, the limitations reported by the studies themselves were consistent. Most concerned scale and design: many studies were small, single-site, single-arm or non-randomised, with short or absent follow-up [12], [30]. Recurring technical and translational obstacles included hardware cost, registration error, latency, computational demand and manual setup, leaving several systems closer to proof-of-concept than workflow-ready tools [51], [54], [59]. Heterogeneous outcomes, positive-result skew and uneven coverage of MR, hybrid AR/VR, diagnostics, pathology, public health and multidisciplinary team use further limited synthesis.

5.6 Implications for future research

These gaps point to a clear research agenda. The field needs larger, controlled and externally validated studies that compare integrated AI-XR systems against meaningful alternatives and report clinical outcomes rather than technical metrics alone. Studies should also report workflow behaviour, set-up burden, clinician acceptance, cost and sustained use. Established reporting standards for early-stage clinical AI [12] and digital-health evidence [19] would make studies more comparable and temper positive-result bias.

There is also room to broaden the technology and the questions asked of it. Mixed reality and generative or conversational AI are likely to feature more prominently, and deserve rigorous evaluation as they mature. The under-represented domains of diagnostics, pathology and public health, along with multidisciplinary team use, are obvious targets for new work. Because integration mode tracked the clinical task, the most productive studies will be those that design the AI-XR connection around the task from the start, rather than bolting immersion onto an AI tool, or analytics onto an immersive one.

Standardised, patient-centred outcomes, cost reporting and clearer disclosure of AI training and validation data would further support comparison, adoption decisions and reproducibility.

5.7 Strengths and limitations of this review

The review has several strengths. It followed PRISMA, spanned the full range of immersive modalities and a broad set of clinical domains rather than a single specialty, and coded each system at two levels: the AI method and the clinical function it served. This kept the technical and clinical readings distinct. The integration taxonomy gave a consistent way to compare otherwise heterogeneous systems, and the synthesis was deliberately careful not to equate positive reports with proven benefit.

Some limitations should temper its conclusions. Several coding decisions, particularly integration mode, evidence maturity and technology readiness level, called for judgement. These were documented to support transparency, but a degree of subjectivity remains. The heterogeneity of designs, outcomes and sample units, together with the strong skew towards positive results, made meta-analysis inappropriate and confined the review to structured narrative and visual synthesis. These constraints bound how far the findings can be generalised, and they are themselves a reminder of how early this field still is.

6 Conclusion

This PRISMA-guided systematic review of 41 clinically relevant empirical studies set out to clarify where and how the integration of AI with AR, VR and MR is approaching real-world impact in medicine. Three findings stand out. First, AI-XR integration has moved beyond proof-of-principle, and is now being evaluated with patients, clinicians and clinical data across a widening set of applications. Second, the way the two technologies connect is governed by the clinical task, not by the AI method family or the display hardware, so the umbrella label AI-XR integration in fact covers structurally distinct architectures. The integration-mode taxonomy used here gives a more faithful basis for comparison than technology-centred groupings. Third, activity should not be mistaken for maturity. Reported outcomes were almost uniformly positive, but the evidence concentrated at small-scale piloting, and no study reached testing at scale or routine adoption.

The implications are clear. Future work should move beyond feasibility through larger, controlled and externally validated studies that report clinical rather than purely technical outcomes, and that treat sustained use within real workflows as an outcome in its own right. Greater transparency about AI components and their training data, attention to under-represented modalities and domains, and clearer arrangements for consent, accountability and equitable access will all be needed. AI-XR is trending towards clinical impact in a few well-suited areas, but closing the gap between promise and durable, deployable benefit, rather than expanding the catalogue of pilots, is the central task for the next phase of this field.

References

[1] World Health Organization, Global strategy on digital health 2020-2025. Geneva, Switzerland: World Health Organization, 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240020924. Accessed: May 31, 2026.

[2] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44-56, Jan. 2019, doi: 10.1038/s41591-018-0300-7, PMID: 30617339.

[3] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," Nature Medicine, vol. 28, pp. 31-38, Jan. 2022, doi: 10.1038/s41591-021-01614-0, PMID: 35058619.

[4] H. Ghaednia et al., "Augmented and virtual reality in spine surgery, current applications and future potentials," The Spine Journal, vol. 21, no. 10, pp. 1617-1625, Oct. 2021, doi: 10.1016/j.spinee.2021.03.018.

[5] B. M. Kyaw et al., "Virtual reality for health professions education: Systematic review and meta-analysis by the Digital Health Education Collaboration," Journal of Medical Internet Research, vol. 21, no. 1, Art. no. e12959, Jan. 2019, doi: 10.2196/12959, PMID: 30668519.

[6] J. Kim, A. S. Campbell, B. E.-F. de Avila, and J. Wang, "Wearable biosensors for healthcare monitoring," Nature Biotechnology, vol. 37, no. 4, pp. 389-406, Apr. 2019, doi: 10.1038/s41587-019-0045-y, PMID: 30804534, PMCID: PMC8183422.

[7] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, Dec. 2017, doi: 10.1016/j.media.2017.07.005, PMID: 28778026.

[8] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," IEICE Transactions on Information and Systems, vol. E77-D, no. 12, pp. 1321-1329, Dec. 1994, doi: 10.1587/e77-d_12_1321.

[9] M. Sharma, C. Savage, M. Nair, I. Larsson, P. Svedberg, and J. M. Nygren, "Artificial intelligence applications in health care practice: Scoping review," Journal of Medical Internet Research, vol. 24, no. 10, Art. no. e40238, Oct. 2022, doi: 10.2196/40238, PMID: 36197712.

[10] B. Debnath, M. O'Brien, M. Yamaguchi, and A. Behera, "A review of computer vision-based approaches for physical rehabilitation and assessment," Multimedia Systems, vol. 28, pp. 209-239, Feb. 2022, doi: 10.1007/s00530-021-00815-4.

[11] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," BMC Medicine, vol. 17, no. 1, Art. no. 195, Oct. 2019, doi: 10.1186/s12916-019-1426-2, PMID: 31665002.

[12] B. Vasey et al., "Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI," Nature Medicine, vol. 28, pp. 924-933, May 2022, doi: 10.1038/s41591-022-01772-9, PMID: 35585198.

[13] K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, and M. Crotty, "Virtual reality for stroke rehabilitation," Cochrane Database of Systematic Reviews, no. 11, CD008349, Nov. 2017, doi: 10.1002/14651858.CD008349.pub4, PMID: 29156493.

[14] W. Deng et al., "The efficacy of virtual reality exposure therapy for PTSD symptoms: A systematic review and meta-analysis," Journal of Affective Disorders, vol. 257, pp. 698-709, Oct. 2019, doi: 10.1016/j.jad.2019.07.086.

[15] D. Freeman et al., "Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial," The Lancet Psychiatry, vol. 5, no. 8, pp. 625-632, Aug. 2018, doi: 10.1016/S2215-0366(18)30226-8, PMID: 30007519.

[16] D. Suresh, A. Aydin, S. James, K. Ahmed, and P. Dasgupta, "The role of augmented reality in surgical training: A systematic review," Surgical Innovation, vol. 30, no. 2, pp. 366-382, Apr. 2023, doi: 10.1177/15533506221140506, PMID: 36412148.

[17] R. Nowak, M. Nowak, M. Rekas, et al., "Intraoperative use of mixed reality (MR) in humans across surgical disciplines: a major review," 3D Printing in Medicine, vol. 11, Art. no. 59, 2025, doi: 10.1186/s41205-025-00305-7.

[18] D. Reiners, M. R. Davahli, W. Karwowski, and C. Cruz-Neira, "The combination of artificial intelligence and extended reality: A systematic review," Frontiers in Virtual Reality, vol. 2, Art. no. 721933, Sep. 2021, doi: 10.3389/frvir.2021.721933.

[19] H. Unsworth et al., ”The NICE Evidence Standards Framework for digital health and care technologies - Developing and maintaining an innovative evidence framework with global impact,” Digital Health, vol. 7, pp. 1-9, Jun. 2021, doi: 10.1177/20552076211018617.

[20] M. J. Page et al., ”The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” BMJ, vol. 372, Art. no. n71, Mar. 2021, doi: 10.1136/bmj.n71, PMID: 33782057.

[21] L. Maier-Hein, S. S. Vedula, S. Speidel, et al., ”Surgical data science for next-generation interventions,” Nature Biomedical Engineering, vol. 1, pp. 691-696, Sep. 2017, doi: 10.1038/s41551-017-0132-7.

[22] A. Tabassum, I. Ghaznavi, A. Abd-Alrazaq, and J. Qadir, ”Exploring the application of AI and extended reality technologies in metaverse-driven mental health solutions: Scoping review,” Journal of Medical Internet Research, vol. 27, Art. no. e72400, Aug. 2025, doi: 10.2196/72400, PMID: 40829151.

[23] E. A. Chance, ”The combined impact of AI and VR on interdisciplinary learning and patient safety in healthcare education: A narrative review,” BMC Medical Education, vol. 25, Art. no. 1039, 2025, doi: 10.1186/s12909-025-07589-7.

[24] Y. Shin, E. Kim, K. Kim, J. S. Kim, S. Kim, and B.-N. Kim, ”Predicting efficacy of virtual reality-based stabilization for individuals with posttraumatic stress symptoms: A machine learning approach,” Psychiatry Investigation, vol. 22, no. 10, pp. 1178-1190, 2025, doi: 10.30773/pi.2025.0168.

[25] M. Robles, N. Namdarian, J. Otto, E. Wassiljew, N. Navab, C. Falter-Wagner, and D. Roth, ”A virtual reality based system for the screening and classification of autism,” IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 5, pp. 2168-2178, May 2022, doi: 10.1109/TVCG.2022.3150489.

[26] J.-H. Park, Y.-B. Shin, D. Jung, J.-W. Hur, S. P. Pack, H.-J. Lee, H. Lee, and C.-H. Cho, ”Machine learning prediction of anxiety symptoms in social anxiety disorder: Utilizing multimodal data from virtual reality sessions,” Frontiers in Psychiatry, vol. 15, Art. no. 1504190, 2025, doi: 10.3389/fpsyt.2024.1504190.

[27] M. Obst et al., ”AI-enhanced 3D models in global virtual reality case conferences for surgical care in a low-income country: Exploratory study,” JMIR Formative Research, vol. 9, Art. no. e69300, 2025, doi: 10.2196/69300.

[28] H. G. Mohamad, M. N. Khan, M. Tahir, N. Ismat, A. Zaffar, F. Naseer, and S. Ali, ”A predictive and adaptive virtual exposure framework for spider fear: A multimodal VR-based behavioral intervention,” Healthcare, vol. 13, no. 20, Art. no. 2617, Oct. 2025, doi: 10.3390/healthcare13202617.

[29] B. Lim and W. Song, ”Development and validation of machine learning models for classifying cancer-related sarcopenia using Kinect-based mixed-reality exercises in breast cancer survivors,” Translational Cancer Research, vol. 14, no. 7, pp. 4208-4218, 2025, doi: 10.21037/tcr-2024-2337.

[30] E. J. Leehr et al., ”Clinical predictors of treatment response towards exposure therapy in virtuo in spider phobia: A machine learning and external cross-validation approach,” Journal of Anxiety Disorders, vol. 83, Art. no. 102448, 2021, doi: 10.1016/j.janxdis.2021.102448.

[31] B. Koo et al., ”Automatic, global registration in laparoscopic liver surgery,” International Journal of Computer Assisted Radiology and Surgery, vol. 17, no. 1, pp. 167-176, Jan. 2022, doi: 10.1007/s11548-021-02518-7.

[32] A. Javanbakht, L. Hinchey, K. Gorski, A. Ballard, L. Ritchie, and A. Amirsadri, ”Unreal that feels real: Artificial intelligence-enhanced augmented reality for treating social and occupational dysfunction in post-traumatic stress disorder and anxiety disorders,” European Journal of Psychotraumatology, vol. 15, no. 1, Art. no. 2418248, 2024, doi: 10.1080/20008066.2024.2418248.

[33] D. Horesh et al., ”Virtual reality combined with artificial intelligence (VR-AI) reduces hot flashes and improves psychological well-being in women with breast and ovarian cancer: A pilot study,” Healthcare, vol. 10, no. 11, Art. no. 2261, Nov. 2022, doi: 10.3390/healthcare10112261.

[34] E. Delamarre, M. Nejjari, J. Dreyfus, F. Lesage, and W. Ben Ali, ”Transcatheter tricuspid valve implantation with LuX-Valve utilizing a novel patient-specific virtual and physical simulator: A case report,” European Heart Journal - Case Reports, vol. 8, no. 11, Art. no. ytae582, Nov. 2024, doi: 10.1093/ehjcr/ytae582.

[35] J. Cui, ”An adaptive hand exoskeleton rehabilitation training system integrating virtual reality and an AI-based assessment engine,” Frontiers in Sports and Active Living, vol. 7, Art. no. 1724021, 2025, doi: 10.3389/fspor.2025.1724021, PMID: 41450855.

[36] D. Amparore et al., ”Indocyanine green drives computer vision based 3D augmented reality robot assisted partial nephrectomy: The beginning of 'automatic' overlapping era,” Urology, vol. 164, pp. e312-e316, Jun. 2022, doi: 10.1016/j.urology.2021.10.053.

[37] M. Al-Rajab et al., ”A smart secure virtual reality immersive application for Alzheimer's and dementia patients,” Scientific Reports, vol. 15, Art. no. 30584, 2025, doi: 10.1038/s41598-025-14624-6, PMID: 40835985.

[38] Y. An, S. Min, and C. Park, ”Mixed reality-based physical therapy in older adults with sarcopenia: Preliminary randomized controlled trial,” JMIR Serious Games, vol. 13, Art. no. e76357, 2025, doi: 10.2196/76357.

[39] X. Huang et al., ”Evaluation of augmented reality surgical navigation in percutaneous endoscopic lumbar discectomy: Clinical study,” Bioengineering, vol. 10, no. 11, Art. no. 1297, Nov. 2023, doi: 10.3390/bioengineering10111297.

[40] E. Lendaro, A. Middleton, S. Brown, and M. Ortiz-Catalan, ”Out of the clinic, into the home: The in-home use of phantom motor execution aided by machine learning and augmented reality for the treatment of phantom limb pain,” Journal of Pain Research, vol. 13, pp. 195-209, 2020, doi: 10.2147/JPR.S220160.

[41] P. Auloge et al., ”Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: A pilot randomised clinical trial,” European Spine Journal, vol. 29, no. 7, pp. 1580-1589, Jul. 2020, doi: 10.1007/s00586-019-06054-6.

[42] S.-H. Lee, J. Cui, L. Liu, M.-C. Su, L. Zheng, and S.-C. Yeh, ”An evidence-based intelligent method for upper-limb motor assessment via a VR training system on stroke rehabilitation,” IEEE Access, vol. 9, pp. 65871-65881, 2021, doi: 10.1109/ACCESS.2021.3075778.

[43] C.-F. Tsai et al., ”A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 2124-2132, 2021, doi: 10.1109/TNSRE.2021.3118918.

[44] A. H. Sadeghi et al., ”Virtual reality and artificial intelligence for 3-dimensional planning of lung segmentectomies,” JTCVS Techniques, vol. 7, pp. 309-321, Jun. 2021, doi: 10.1016/j.xjtc.2021.03.016.

[45] L. Cai et al., ”Improving Ki67 assessment concordance by the use of an artificial intelligence-empowered microscope: A multi-institutional ring study,” Histopathology, vol. 79, no. 4, pp. 544-555, Oct. 2021, doi: 10.1111/his.14383, PMID: 33840132.

[46] D. C. Beidel, P. W. Tuerk, J. Spitalnick, C. A. Bowers, and K. Morrison, ”Treating childhood social anxiety disorder with virtual environments and serious games: A randomized trial,” Behavior Therapy, vol. 52, no. 6, pp. 1351-1363, Nov. 2021, doi: 10.1016/j.beth.2021.03.003.

[47] Y. Bai, F. Liu, and H. Zhang, ”Artificial intelligence limb rehabilitation system on account of virtual reality technology on long-term health management of stroke patients in the context of the internet,” Computational and Mathematical Methods in Medicine, vol. 2022, Art. no. 2688003, 2022, doi: 10.1155/2022/2688003.

[48] R. L. Harms et al., ”Digital biomarkers and sex impacts in Alzheimer’s disease management - potential utility for innovative 3P medicine approach,” EPMA Journal, vol. 13, no. 2, pp. 311-322, Jun. 2022, doi: 10.1007/s13167-022-00284-3.

[49] A. Hudon, M. Beaudoin, K. Phraxayavong, S. Potvin, and A. Dumais, ”Unsupervised machine learning driven analysis of verbatims of treatment-resistant schizophrenia patients having followed avatar therapy,” Journal of Personalized Medicine, vol. 13, no. 5, Art. no. 801, May 2023, doi: 10.3390/jpm13050801.

[50] S.-Y. Chiou et al., ”Augmented reality surgical navigation system integrated with deep learning,” Bioengineering, vol. 10, no. 5, Art. no. 617, May 2023, doi: 10.3390/bioengineering10050617.

[51] S. L. Metzger et al., ”A high-performance neuroprosthesis for speech decoding and avatar control,” Nature, vol. 620, no. 7976, pp. 1037-1046, Aug. 2023, doi: 10.1038/s41586-023-06443-4.

[52] E. Checcucci et al., ”Three-dimensional automatic artificial intelligence driven augmented-reality selective biopsy during nerve-sparing robot-assisted radical prostatectomy: A feasibility and accuracy study,” Asian Journal of Urology, vol. 10, no. 4, pp. 407-415, Oct. 2023, doi: 10.1016/j.ajur.2023.08.001.

[53] A. Hudon, M. Beaudoin, K. Phraxayavong, S. Potvin, and A. Dumais, ”Enhancing predictive power: Integrating a linear support vector classifier with logistic regression for patient outcome prognosis in virtual reality therapy for treatment-resistant schizophrenia,” Journal of Personalized Medicine, vol. 13, no. 12, Art. no. 1660, Nov. 2023, doi: 10.3390/jpm13121660.

[54] D. Amparore et al., ”Computer vision and machine-learning techniques for automatic 3D virtual images overlapping during augmented reality guided robotic partial nephrectomy,” Technology in Cancer Research & Treatment, vol. 23, Art. no. 15330338241229368, 2024, doi: 10.1177/15330338241229368.

[55] S. Oh et al., ”Diagnosis of ADHD using virtual reality and artificial intelligence: An exploratory study of clinical applications,” Frontiers in Psychiatry, vol. 15, Art. no. 1383547, 2024, doi: 10.3389/fpsyt.2024.1383547.

[56] Y. C. Kim, C.-U. Park, S. J. Lee, W. S. Jeong, S. W. Na, and J. W. Choi, ”Application of augmented reality using automatic markerless registration for facial plastic and reconstructive surgery,” Journal of Cranio-Maxillofacial Surgery, vol. 52, no. 2, pp. 246-251, Feb. 2024, doi: 10.1016/j.jcms.2023.12.009.

[57] L. Casanova-Lozano, D. Reifs-Jimenez, M. Marti-Ejarque, R. Reig-Bolano, and S. Grau-Carrion, ”Evaluation of two digital wound area measurement methods using a non-randomized, single-center, controlled clinical trial,” Electronics, vol. 13, no. 12, Art. no. 2390, Jun. 2024, doi: 10.3390/electronics13122390.

[58] X. Hou, X. Liao, R. Xu, F. Fei, and B. Wu, ”Open-source AI-assisted rapid 3D color multimodal image fusion and preoperative augmented reality planning of extracerebral tumors,” Neurosurgical Focus, vol. 59, no. 1, Art. no. E12, Jul. 2025, doi: 10.3171/2025.4.FOCUS24557.

[59] F. Porpiglia et al., ”Artificial intelligence 3D augmented reality-guided robotic prostatectomy versus cognitive MRI intervention: Results of the prospective randomized RIDERS trial,” European Urology, early access, 2025, doi: 10.1016/j.eururo.2025.09.4172.

[60] M. Bulduk, V. Can, E. Aktas, B. Ipekci, B. Bulduk, and I. Nas, ”Artificial intelligence-assisted virtual reality for reducing anxiety in pediatric endoscopy,” Journal of Clinical Medicine, vol. 14, no. 4, Art. no. 1344, Feb. 2025, doi: 10.3390/jcm14041344.

[61] E. Lendaro et al., ”Extended reality used in the treatment of phantom limb pain: A multicenter, double-blind, randomized controlled trial,” Pain, vol. 166, no. 3, pp. 571-586, Mar. 2025, doi: 10.1097/j.pain.0000000000003384.

[62] J. Wang et al., ”An adaptive AI-based virtual reality sports system for adolescents with excess body weight: A randomized controlled trial,” Nature Medicine, vol. 31, no. 7, pp. 2255-2268, Jul. 2025, doi: 10.1038/s41591-025-03724-5.

[63] S. Wang, Y. Zhang, F. Xu, and G. Zhou, ”Intelligent monitoring system for quality of life of colostomy patients based on deep learning and AR,” Scientific Reports, vol. 15, Art. no. 37507, 2025, doi: 10.1038/s41598-025-21470-z.

[64] M. Munoz-Novoa, M. B. Kristoffersen, K. S. Sunnerhagen, A. Naber, M. Ortiz-Catalan, and M. Alt Murphy, ”Myoelectric pattern recognition with virtual reality and serious gaming improves upper limb function in chronic stroke: A single case experimental design study,” Journal of NeuroEngineering and Rehabilitation, vol. 22, no. 1, Art. no. 6, Jan. 2025, doi: 10.1186/s12984-025-01541-y.

[65] Life Sciences Hub Wales, Understanding Sector-Based Technology Readiness Levels: Pharmaceuticals, Medical Devices, Digital and Pathways. Cardiff, Wales:

Life Sciences Hub Wales, Jun. 2025. [Online]. Available: https://lshubwales.com/sites/default/files/2025-08/202506%20TRL%20advanced%20-%20eng.pdf. Accessed: May 31, 2026.

[66] NSW Health, "Technology readiness level (TRL) development pathway," NSW Government, May 20, 2024. [Online]. Available: https://www.health.nsw.gov.au/services/Pages/translational-research-pathway.aspx. Accessed: May 31, 2026.