Augmented-Reality Based Assistive Systems for Critical Medical Procedures

Dissertation

Thesis approved by the Department of Computer Science
RPTU University Kaiserslautern-Landau
for the award of the Doctoral Degree
Doctor of Engineering (Dr.-Ing.)

to

Hamraz Javaheri

Date of Defense: December 17, 2025
Dean: Prof. Dr. Christoph Garth
Reviewer: Prof. Dr. Paul Lukowicz
Reviewer: Prof. Dr. Thad Starner

DE-386

Zusammenfassung

Augmented Reality (AR) birgt ein erhebliches Potenzial zur Steigerung der menschlichen Leistungsfähigkeit in risikoreichen, sicherheitskritischen Bereichen, in denen Fehler schwerwiegende Folgen haben können. Durch die direkte Einblendung digitaler Informationen in das Sichtfeld des Benutzers kann AR die schnelle Entscheidungsfindung unterstützen, das Situationsbewusstsein aufrechterhalten und die Präzision unter kritischen Bedingungen verbessern. Die Komplexität solcher Bereiche und der begrenzte Umfang der interdisziplinären Forschung haben jedoch die Entwicklung geeigneter AR-basierter Systeme für diese spezifischen Anwendungsfälle behindert.

Diese Dissertation untersucht, wie AR genutzt werden kann, um die menschliche Leistungsfähigkeit bei Aufgaben zu verbessern, die erhebliche kognitive Anstrengungen, präzise motorische Kontrolle und schnelle, komplexe Entscheidungen unter Druck erfordern. Kritische medizinische Verfahren, bei denen diese Herausforderungen von Natur aus vorhanden sind, dienen als primärer Untersuchungskontext. Durch einen interdisziplinären Ansatz, der Informatik und medizinische Praxis integriert, entwickelt die Arbeit Methoden zur Unterstützung von Klinikern in anspruchsvollen Umgebungen.

Zu diesem Zweck konzentriert sich diese Arbeit zunächst auf den Aufbau von Grundlagenwissen über die Anforderungen an funktionsfähige AR-Systeme, die die menschliche Leistungsfähigkeit verbessern können. Anschließend untersucht sie den Gestaltungsraum und die Auswirkungen von AR auf den Erwerb von Verfahrens- und Motorikfähigkeiten, indem sie die Ausbildung von Krankenschwestern in einem Szenario der Herz-Lungen-Wiederbelebung untersucht. Danach untersucht sie AR-Systeme und ihre Rolle bei der Entscheidungsfindung und kognitiven Unterstützung während risikoreicher Eingriffe, wobei ein besonderer Schwerpunkt auf ihrer Integration in die Operation zur Entfernung von Bauchspeicheldrüsentumoren liegt. Anhand klinischer Studien werden darüber hinaus die nachgelagerten, realen Auswirkungen solcher Systeme auf die chirurgische Leistung und die Patientenergebnisse untersucht. Schließlich werden kontextsensitive AR-Designs unter Verwendung großer Sprachmodelle untersucht, um zusätzliche kognitive Unterstützung und Entscheidungsfähigkeiten zu bieten. Auf der Grundlage all dieser Untersuchungen werden in der Arbeit Designimplikationen für die zukünftige Entwicklung von AR-Systemen in medizinischen Umgebungen mit hohem Risiko abgeleitet.

Abstract

Augmented reality (AR) holds significant potential to enhance human performance in high-stakes, safety-critical domains where errors can have severe consequences. By overlaying digital information directly into the user's field of view, AR can support rapid decision-making, maintain situational awareness, and improve precision under critical conditions. However, the complexity of such domains and the limited scope of interdisciplinary research have hindered the development of suitable AR-based systems for these specific use cases.

This dissertation investigates how AR can be harnessed to improve human performance in tasks that require substantial cognitive effort, precise motor control, and rapid, complex decision-making under pressure. Critical medical procedures, in which these challenges are inherently present, serve as the primary context for investigation. Through an interdisciplinary approach that integrates computer science and medical practice, the work develops methods designed to support clinicians in high-demand environments.

To this end, this thesis first focuses on building foundational knowledge about the requirements for viable AR systems that can enhance human performance. It then explores the design space and impact of AR for procedural and motor skill acquisition by examining nurse training in a cardiopulmonary resuscitation scenario. Following this, it investigates AR systems and their role in decision-making and cognitive support during high-risk procedures, with a specific focus on their integration into pancreatic tumor removal surgery. Through clinical trials, it further examines the downstream, real-world impacts of such systems on surgical performance and patient outcomes. Finally, it explores context-aware AR designs using large language models to provide additional cognitive support and decision-making capabilities. Drawing on all of these investigations, the thesis distills design implications for the future development of AR systems in high-stakes medical environments.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Paul Lukowicz, for his year-long support and for giving me the opportunity to follow my dream research path. I am especially grateful to Dr. Stavrou for his amazing, unlimited, and kind support throughout my thesis, and for providing me with opportunities to explore interdisciplinary research directions that would otherwise not have been feasible. My heartfelt thanks also go to my mentor, Dr. Jakob Karolus, who taught me the ways of doing research and supported me through every stage of my thesis journey.

I am thankful to my project collaborators, Eloise Monger and Mary Gobby from the University of Southampton, with whom it has always been a joy and a pleasure to work. I also wish to thank my dear colleagues, Bo, Vitor, Daniel, Mengxi, Joanna, Shannon, Jane, Hymalai, Passant, Max, Lars, and Lala, for being the best colleagues in the world. A special thanks goes to my colleague and friend Agnes, who made all of our occasions and especially our trips, unforgettable for me and filled them with joyful memories.

Last, but most definitely not least, I want to give my most heartfelt thanks to my family. To my parents, who have stood by me through every step of my journey, always having my back through ups and downs. To my husband Omid, my research companion, my mentor, and my best friend in the entire world, whom I could never imagine a single moment without. To my siblings, Raha, Reza, Aida, and Faraz, who have been there for me throughout every phase of my life. And to my beloved babies, Loki, Lily, Charlie, Pablo, Frodo, and Caesar, who have filled every day of my life with joy and happiness, and I could never imagine a life without them.

December 17, 2025, Hamraz Javaheri

To Dr. med. Omid Ghamarnejad – A surgeon, a partner, and my greatest source of strength. May this work inspire others to embrace their differences, as we have, to create something greater together.

Contents

List of Figures

Chapter 1

Introduction

1.1. Motivation

The emergence of technologies such as Augmented Reality (AR), artificial intelligence (AI), and wearable computing has introduced new opportunities for enhancing clinical decision-making, procedural accuracy, and medical training. Among these, wearable AR systems, particularly head-mounted displays (HMDs), have shown substantial promise in supporting complex, high-risk, and time-sensitive medical procedures. By delivering real-time, context-aware visual information directly into the clinician's field of view, these systems offer hands-free interaction, intraoperative guidance, and enhanced situational awareness—features that may reduce cognitive load and ultimately improve patient outcomes.

Despite their potential, the clinical adoption of AR remains limited, especially in acute and high-demand environments such as emergency medicine, surgery, and critical care. One of the core issues is the disconnect between academic innovation and real-world clinical applicability. Many AR applications remain confined to controlled laboratory settings or proof-of-concept stages, with minimal validation in actual clinical workflows. This results in systems that may be technically functional but lack the usability, reliability, or safety required in demanding healthcare settings.

Another major challenge is the lack of attention to human factors in AR system design. Much of the existing research focuses on technical feasibility, while overlooking critical elements such as user dexterity, ergonomic comfort, cognitive load, and compatibility with fast-paced medical workflows. In cognitively and/or physically demanding scenarios, even minor design oversights can significantly affect user performance and patient safety. Furthermore, the absence of standardized evaluation frameworks, design principles, and best practices has contributed to fragmented, non-generalizable findings across the field.

Compounding these challenges is the dynamic and unpredictable nature of high-demand clinical procedures, which demand AR systems that are not only technically robust but also adaptive and context-aware. While recent advance-

ments in artificial intelligence, such as Large Language Models (LLMs), offer new possibilities for intelligent interaction, these capabilities remain largely unexplored in AR-based clinical applications. The opportunity to reduce cognitive burden and enable seamless, context-sensitive assistance through such technologies is substantial, yet under-investigated.

To address these gaps, this dissertation employs a structured, three-phase research approach aimed at identifying barriers and uncovering viable pathways for the effective integration of AR into medical practice. The first phase begins with a comprehensive literature review that synthesizes prior work, identifies applicable methodologies, and highlights areas for improvement. Building on this foundation, the research evaluates the impact of hardware- and software-related factors on user performance and system usability. This phase builds a foundational body of domain-specific knowledge that clarifies the prerequisites, capabilities, and limitations of current AR technologies for use in healthcare settings.

The second phase of the thesis operationalizes these insights through the design, engineering, and real-world evaluation of two representative use case scenarios: one focused on AR for procedural and motor skill acquisition, and the other on AR as a tool for cognitive support and clinical decision-making. To this end, two AR systems were developed. The first, RescuAR, is a self-guided AR training system for cardiopulmonary resuscitation (CPR) that delivers real-time, multimodal feedback to enhance procedural and motor skill acquisition among both laypeople and nurse students. The second, ARAS, is an AR assistance system tailored for open pancreatic surgery, supporting preoperative planning and in situ anatomical visualization. These two use scenarios were intentionally selected to reflect a continuum of complexity and clinical criticality. This progression from training to assistance enables a comprehensive exploration of AR's role across diverse demanding clinical scenarios and user needs. Additionally, this phase introduces a novel interaction paradigm to reduce task-induced cognitive load, leveraging LLMs to enable natural language communication and context-aware system behavior while simplifying system design. Through a series of user studies, including a clinical trial, all designed and developed systems were rigorously evaluated using a combination of quantitative performance metrics and qualitative user feedback. These evaluations demonstrate not only the feasibility of the proposed approaches but also their real-world clinical impact. In particular, the ARAS use case shows that AR systems tangibly improved both surgical performance and patient outcomes, as evidenced by a matched-pair analysis study.

In the final phase, the dissertation reflects on qualitative insights and practical lessons learned throughout the research to contribute to domain-specific human-computer interaction (HCI) knowledge. This includes an exploration of the broader, multifaceted considerations, such as safety, usability, ethics, and interdisciplinary collaboration, critical to developing AR systems for high-stakes medical environments.

By integrating perspectives from system engineering, computer science, HCI, and clinical practice, this research advances our understanding of how AR tech-

nologies can be effectively designed, validated, and translated into practical, user-centered tools that assist users during highly demanding medical scenarios. Through a strong emphasis on clinical relevance and real-world evaluation, the dissertation not only contributes to theoretical discourse but also provides actionable insights and design guidelines to support the responsible and effective adoption of AR in healthcare. Ultimately, the findings inform future AR system design and deployment strategies aimed at improving the quality, safety, and efficiency of patient care in time-critical medical scenarios.

1.2. Objectives of the Study

The overarching aim of this dissertation is to investigate the potential of AR to enhance human performance in tasks characterized by high cognitive load, fine motor precision, and complex decision-making under pressure. Critical medical procedures, which inherently involve these demands, serve as the primary application context for this research. By adopting an interdisciplinary approach that bridges computer science and medical practice, this work focuses on developing methods to support clinicians operating in such environments. To guide this investigation, the dissertation is structured around the following five hypotheses:

1. **Device and Modality Effects on Performance:** The choice of device type and interaction modality in wearable AR systems significantly influences human performance metrics.

2. **AR for Procedural and Motor Skill Learning:** AR-based systems that provide real-time, multimodal feedback improve motor and procedural skill acquisition more effectively than traditional interactive training methods in the medical domain.

3. **AR for Cognitive Load Reduction:** The use of AR for task guidance during cognitively demanding clinical procedures reduces perceived cognitive load and improves user confidence and task efficiency compared to non-AR alternatives.

4. **Downstream Clinical Impact of AR Assistance:** The integration of AR-based assistive tools into surgical workflows leads to measurable downstream improvements in clinical outcomes, including intraoperative and patient recovery metrics.

5. **Context-Aware, Natural Interaction for Cognitive Support:** The integration of context-aware system behavior and natural communication (e.g., via LLMs) in AR interfaces reduces task-induced cognitive load and improves interaction fluency in complex clinical workflows.

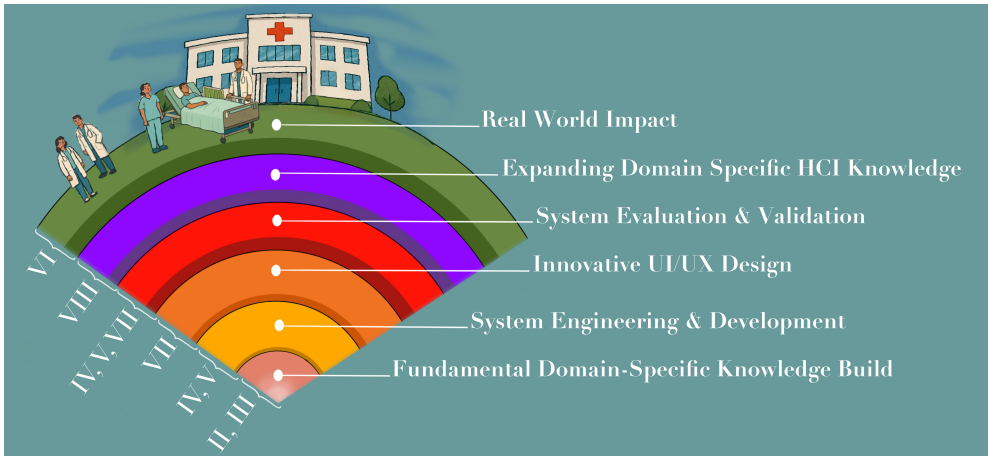


Figure 1.1.: Thesis contribution overview and associated chapter numbers.

1.3. Contributions

This thesis makes six major contributions (Figure 1.1) that collectively advance the technical, clinical, and human-centered dimensions of medical AR research:

1.3.1. Foundational Domain-Specific Knowledge Build

This dissertation presents a comprehensive investigation that integrates user-centered studies, literature review, and empirical evaluations to examine how hardware- and software-related factors affect human performance and perceived system usability. It identifies critical human and technological constraints that inform the design, selection, and deployment of AR systems in the medical domain. As a result, this work builds foundational domain-specific knowledge, offering empirical evidence and design insights that are essential for guiding the future development of medical AR hardware and interaction modalities.

1.3.2. System Engineering and Development

This dissertation encompasses two distinct AR applications that span a continuum of complexity and clinical criticality, illustrating a progression of increasing technical, ergonomic, and domain-specific challenges. These use cases serve as anchors for demonstrating robust system engineering and user-centered design approaches across diverse medical contexts, one focused on procedural and motor skill acquisition in training scenarios, and the other on cognitive support during high-stakes surgical procedures:

- **RescuAR:** A self-guided AR training platform for cardiopulmonary resuscitation (CPR), developed and validated through controlled quantitative studies to evaluate training efficacy, usability, and user performance.

- **ARAS:** A clinically integrated AR system for preoperative planning and intraoperative assistance in open pancreatic surgery. Representing a significantly higher degree of complexity, ARAS required real-time, *in situ* anatomical visualization, seamless integration with surgical workflows, and precise spatial registration, addressing both technical and clinical constraints inherent to high-risk surgical environments.

Insights derived from the engineering and design of these systems offer transferable methodologies and guiding principles that can inform the development of future AR solutions across a broad range of medical domains and clinical scenarios.

1.3.3. Innovative User Experience and User Interface Design

As future interactive systems increasingly move toward intelligent and personalized design, this thesis explores novel strategies for enhancing user experience and interface adaptability through the integration of LLMs in high-stakes clinical contexts. It demonstrates not only the potential of LLMs to reduce users' cognitive load and task completion time, but also their ability to simplify the development process by dynamically combining basic system functionalities into more complex, context-aware behaviors. This work provides early evidence of the feasibility and utility of LLM-driven interaction in critical domains, offering a new perspective on adaptive, user-centered medical AR systems.

1.3.4. System Evaluation and Validation

Beyond offering insights into the engineering and design of the aforementioned systems, this thesis rigorously evaluates their effectiveness through empirical studies involving end users from the medical domain, including clinicians and nursing students. By conducting user studies with these target populations and deploying the systems in real-world settings, including clinical trials, it advances beyond controlled laboratory experiments to establish the ecological validity of the proposed methods and designs. These in-situ evaluations not only validate system usability and effectiveness in high-stakes environments but also demonstrate the practical feasibility of integrating such technologies into existing clinical workflows.

1.3.5. Expanding Domain-Specific HCI Knowledge

This thesis advances domain-specific human-computer interaction (HCI) knowledge by critically examining the broader socio-technical dimensions of designing AR systems for high-stakes medical environments. Building on the technical foundations laid in earlier chapters, it synthesizes insights from interdisciplinary collaboration, qualitative user feedback, and HCI literature to address challenges that extend beyond engineering. Through four key lenses-

co-creation across disciplines, human-centered design and user experience, clinical workflow integration, and domain-specific generalizability- this work articulates a holistic framework for designing effective, trustworthy, and context-aware AR systems in healthcare. In doing so, it contributes conceptual and methodological guidance for future HCI research in safety-critical and highly regulated domains.

1.3.6. Real-World Impact

A key contribution of this thesis lies in demonstrating the real-world impact of AR technologies through their deployment and evaluation in clinical practice. Unlike many studies whose findings remain confined to controlled laboratory environments, this research shows how the developed AR systems tangibly improved both surgeon and patient experiences. Specifically, the integration of AR into intraoperative workflows enhanced surgical precision, spatial awareness, and decision-making efficiency, leading to measurable improvements in surgical performance. These upstream improvements translated into downstream clinical benefits, including reduced operative time and improved postoperative recovery outcomes. These results underscore the practical relevance and translational value of this work, illustrating how well-designed AR solutions can directly benefit patient care. The findings not only validate the proposed systems in ecologically valid settings but also serve to encourage future research focused on bridging the gap between technical innovation and measurable clinical impact.

1.4. Structure of the Thesis

This dissertation is organized into nine chapters that progressively develop the investigation into wearable AR assistance systems for time-critical medical procedures (Figure 1.2). Following this introductory chapter, Chapter 2 presents a comprehensive review of related work, focusing on AR applications in medical education and surgical domains, and highlights the research gaps motivating this study.

Chapter 3 examines the prerequisites for integrating AR technologies in medical environments, including the feasibility of head-mounted displays for precision tasks and their impact on user dexterity and cognitive load. Chapters 4, 5, and 6 then detail two representative use case scenarios. Chapter 4 introduces *RescuAR*, a self-directed AR system designed for cardiopulmonary resuscitation (CPR) training, discussing its design, implementation, and training efficacy. Chapter 5 focuses on system design, development, and evaluation of *ARAS*, an AR-based surgical assistance and navigation system developed for open pancreatic surgery. Following this, Chapter 6 presents real-world outcomes of ARAS during clinical trials by conducting a matched-pair analysis comparing patient and surgical outcomes of patients undergoing surgeries with ARAS and without.

In Chapter 7, the dissertation explores the incorporation of intelligent technologies such as LLMs to enable context-aware AR assistance, enhancing decision support and cognitive load management in dynamic medical settings. Building upon findings presented in previous chapters and drawing on insights from the broader fields of Human-Computer Interaction (HCI), Chapter 8 takes a step beyond the engineering perspective. It shifts focus toward the broader, multifaceted considerations essential for designing and developing systems within the high-stakes context of the medical domain.

Finally, Chapter 9 concludes the dissertation by summarizing the main findings, contributions, and implications for the development of AR systems to support healthcare professionals in time-sensitive procedures and outlining potential directions for future work in this evolving field.

This structured approach guides the reader through the theoretical foundations, system development, practical evaluations, and broader reflections essential to advancing AR-assisted medical applications.

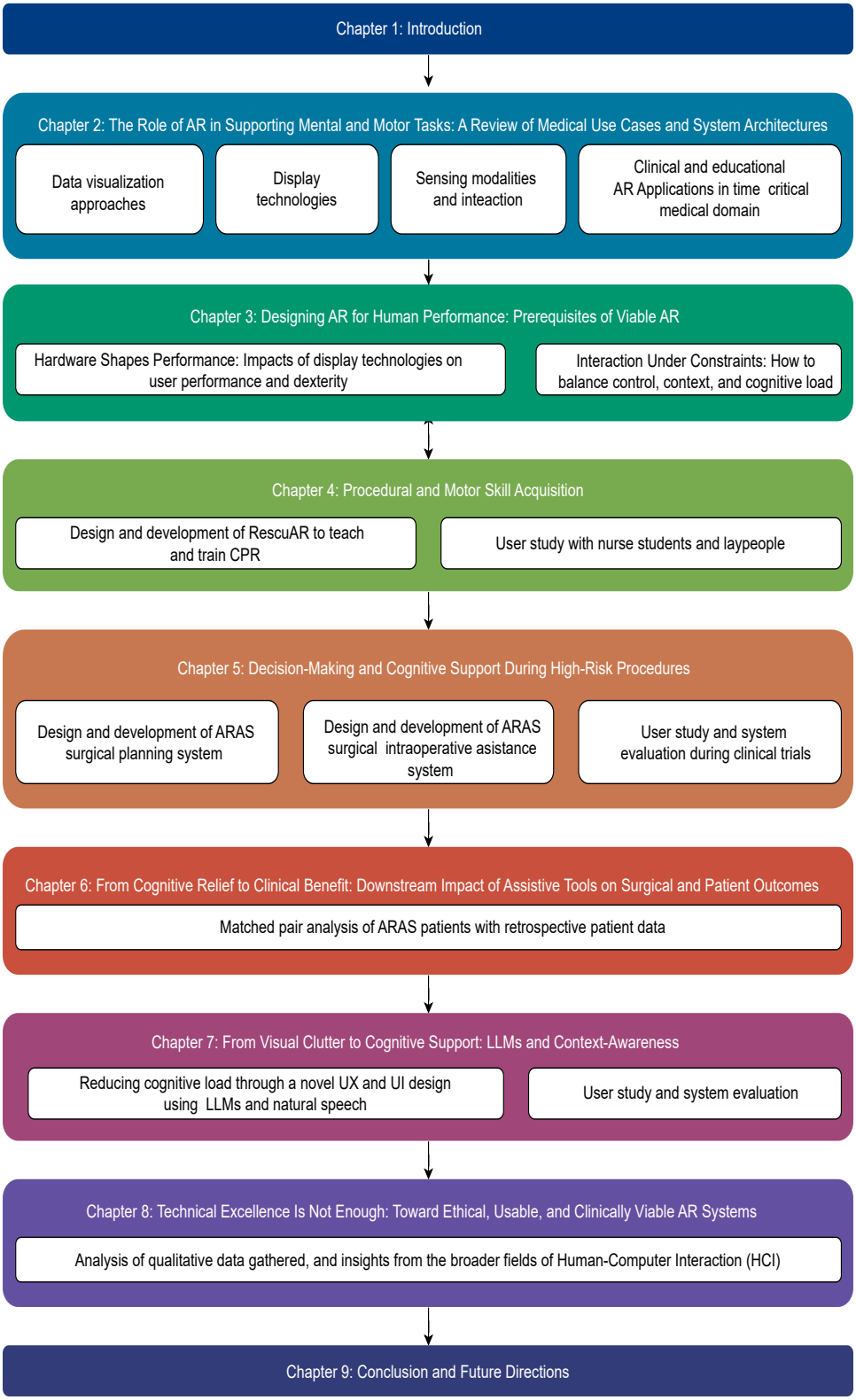


Figure 1.2.: Thesis outline

Chapter 2

The Role of AR in Supporting Mental and Motor Tasks: A Review of Medical Use Cases and System Architectures

Contents

This chapter presents a detailed review of the current state-of-the-art in the use of AR, with a particular focus on its usage in time-critical medical applications. By overlaying digital content onto the real world, AR supports medical learners and professionals in various ways, particularly valuable in time-sensitive and critical domains where rapid decision-making and procedural accuracy are essential.

To identify relevant studies and technologies within this field, a literature search was conducted. The goal was to map the current research landscape, evaluate the effectiveness of AR applications, and identify gaps and challenges in the implementation of AR for acute medical use cases. The search was carried out across four major academic databases: PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar.

2.1. Technical Background

The technical architecture of AR systems in healthcare is defined by a convergence of display modalities, visualization methods, tracking technologies, registration algorithms, and interaction modules. These foundational components determine the usability, accuracy, and effectiveness of AR in clinical environments.

2.1.1. Data Visualization

In-situ Visualization

In-situ AR systems are designed to align virtual information directly onto real anatomical structures. Fritz et al. [90] and Kersten-Oertel et al. [133] demonstrated how in-situ AR in image-guided neurosurgery enhances procedural navigation by reducing the cognitive effort of mentally mapping imaging data. These systems typically rely on optical tracking and real-time registration pipelines to maintain anatomical coherence as the surgical field or camera moves.

Accurate registration is fundamental to the performance of In-situ AR systems in healthcare. It ensures the correct alignment between virtual models and the physical patient, thus maintaining spatial coherence. Multiple registration strategies have been developed, each with its advantages and constraints depending on the clinical context and system requirements.

Manual Registration Manual Registration involves user-defined alignment, typically through visual feedback or alignment of anatomical landmarks. While simple, it is prone to variability and relies heavily on operator expertise [236].

Point-based Registration Point-Based Registration is one of the most widely used methods, where specific anatomical or fiducial points in physical space are mapped to corresponding virtual coordinates. Zheng et al. [283] provide a comprehensive review, highlighting its effectiveness in rigid structures like bones but also its susceptibility to point localization error.

Surface Registration Surface-Based Registration methods use full or partial anatomical surfaces captured via 3D scanning or stereo endoscopy. These approaches reduce reliance on individual points, providing higher robustness and accuracy in soft tissue contexts. Maier-Hein et al. [166] explored this method extensively in minimally invasive surgery using optical surface reconstruction techniques.

Marker-based Registration Marker-Based Registration utilizes fiducial markers (e.g., QR codes, infrared-reflective dots) placed on or near the anatomy. These markers are easy to detect and allow for automated, real-time tracking.

Sielhorst et al. [236] discuss how such methods became foundational in surgical AR, though they require marker placement and are prone to occlusion issues.

Calibration-based Registration Calibration-based registration links AR content to imaging systems (like C-arms) through geometric calibration. Navab et al. [189] introduced the Camera Augmented Mobile C-arm (CAMC) system, which used internal calibration parameters to enable AR overlays that move accurately with the imaging device.

Advanced tracking methods, such as SLAM (Simultaneous Localization and Mapping), continue to evolve in medical AR. Klein and Murray [141] developed one of the early SLAM frameworks for AR, supporting dynamic environments without the need for external markers, an innovation now seen in modern head-mounted displays.

Collectively, these methods form the technical backbone of medical AR, with ongoing research focusing on hybrid approaches that fuse multiple registration techniques to overcome anatomical variability, motion artifacts, and occlusion.

Ex-situ Visualization

Ex-situ AR systems have seen widespread adoption in surgical contexts due to their relative simplicity, compatibility with existing workflows, and ease of implementation. Unlike in-situ systems that overlay data directly onto the patient's anatomy, ex-situ AR displays digital content on separate screens or mobile devices, requiring the surgeon to mentally integrate the virtual and real spaces. Navab et al. [188] explored these trade-offs in operating room deployments, noting that although ex-situ systems can increase cognitive demand, they remain more accessible in resource-constrained settings. Louis et al. [161] demonstrated the use of synchronized virtual and augmented platforms to assist preoperative planning and intraoperative referencing in neurosurgery using head-tracked displays rather than overlaid visualizations. Similarly, Kim et al. [136] developed a monitor-based AR system for collaborative discussion in cardiology, enabling clinicians to manipulate and annotate 3D anatomical models in real time without requiring head-mounted displays or in-field projection.

These systems also enable a greater degree of collaboration and documentation. For instance, Bui et al. [34] highlighted the role of non-in-situ AR in surgical rehearsal and education, where trainees could manipulate anatomical models on screen using spatial controllers. Although these platforms lack the direct spatial congruence of in-situ AR, they provide flexibility, remote accessibility, and easier integration with PACS systems and surgical video feeds. Consequently, non-in-situ AR remains a viable and scalable solution for surgical training, planning, and consultation, particularly in institutions that prioritize interoperability and cost-effectiveness.

2.1.2. Display Technologies

AR display technologies have been widely adopted across various healthcare domains, with device selection often dictated by clinical context, required immersion, and interaction complexity. Optical See-Through Head-Mounted Displays (OST-HMDs), such as the Microsoft HoloLens, are the most frequently used devices in surgery and training due to their ergonomic design and real-time overlay capability. Birlo et al. [24] and Doughty et al. [65] reviewed OST-HMD applications in surgery, highlighting improvements in spatial accuracy, surgical field awareness, and hands-free control. Ahmad et al. [3] demonstrated HoloLens-based visualization in dental implantology, improving procedural guidance and spatial planning.

Video See-Through (VST) HMDs are less common but preferred in simulation and telemedicine due to their complete control over the visual scene. These systems have shown promise in training scenarios where realism and controlled augmentation are prioritized [3]. Light-field and holographic displays are emerging as next-generation visualization tools. Tamboli et al. [251] explored their potential for deep parallax rendering and collaborative holography in preoperative planning, albeit with current limitations in hardware cost and system complexity.

Handheld AR platforms, including tablets and smartphones, offer accessible alternatives for education, triage, and patient education. Oun et al. [202] compared mobile-based AR with HMDs and found that while handheld devices lack spatial immersion, they offer superior portability and are better suited for quick deployments in resource-constrained environments.

2.1.3. Sensing Modalities and Interaction

Modern AR systems in healthcare increasingly rely on diverse sensing modalities and natural interaction methods to enhance user experience, reduce friction, and improve context-awareness. These modalities include touch, voice, gesture, gaze tracking, biosensors, and haptic feedback, each offering unique affordances for specific clinical and training scenarios.

Gesture recognition and hand tracking are among the most common interfaces in head-mounted AR displays, enabling sterile and touchless interaction with virtual content. Papadopoulos et al. [205] provided an extensive overview of AR interaction types, categorizing gesture-based, voice-based, and eye-based controls for immersive environments. Gaze-based interaction is gaining traction for its ability to provide intuitive control, reduce hand burden, and increase cognitive alignment; Plopski et al. [211] detailed gaze tracking as a powerful input modality in surgical and assistive AR applications.

Voice commands offer a hands-free solution, particularly useful in surgical and high-hygiene settings. Sumak et al. [247] emphasized the role of speech recognition and biosignal monitoring for enabling intelligent systems capable of context-sensitive responses. Biosensors, such as heart rate and skin conductance sensors, enable AR systems to adapt content or alert thresholds based on physiological feedback, providing adaptive and patient-specific interfaces

[281].

Touch and haptic modalities also contribute to richer feedback systems. Kim et al. [137] reviewed wearable sensor integration with haptic gloves and vests, showing their capacity to provide force feedback during medical simulation and training tasks. This physical realism augments the user's sensory immersion, helping bridge the gap between simulation and real-world procedures.

Furthermore, sensor fusion and integration further enhance AR platforms by improving interaction fidelity and situational awareness across dynamic clinical environments. Modern AR systems increasingly incorporate wearable biosensors, inertial measurement units (IMUs), depth cameras, and eye-tracking modules to enrich user input and environmental mapping. For instance, Cheng et al. [44] highlight how IMUs and photometric depth sensors can improve spatial tracking and environmental understanding in AR-assisted procedures. Similarly, Khan et al. [134] discuss the integration of wearable biosensors with AR for continuous physiological monitoring, enabling real-time overlay of patient data during intervention.

In addition, gaze-based interaction is gaining traction in head-mounted displays, particularly in optical see-through systems. Eye-tracking enables gaze-directed navigation and context-sensitive annotation, which reduces the cognitive load associated with manual control [210]. These systems can anticipate clinician focus, thereby enhancing responsiveness and minimizing the need for direct input. Underlying these technologies are motion-tracking techniques detailed by Moeslund et al. [184], who describe the evolution of vision-based human motion analysis essential to multimodal AR interfaces.

Despite technical advancements, challenges persist. Occlusion handling, where virtual objects must appear behind real anatomy, and calibration drift are common pitfalls in prolonged use. Moreover, maintaining low-latency rendering and stable hologram anchoring across diverse clinical lighting and motion conditions remains an active area of development.

The trajectory of AR technology in healthcare suggests increasing convergence between computer vision, wearable computing, and real-time surgical data. These technologies are now moving toward context-aware systems that anticipate user intent and environmental changes, marking a shift from tools to intelligent collaborators in clinical care.

2.2. Use Cases

Although AR has existed for several decades [107], recent advancements in visual technologies and the development of novel AR applications have significantly increased interest among both consumers and professionals. In the medical field, AR has seen widespread adoption, benefiting patients, caregivers, and healthcare professionals alike [32, 130].

For patients, AR applications have been utilized in various domains, including patient education [259], rehabilitation [206], distraction therapy [32], and pain and stress management [155, 171]. For caregivers and medical professionals, AR has proven valuable in areas such as education and training [6,

39, 185, 187], as well as in clinical settings where it serves as a supportive tool [15, 87, 277].

Given the scope of this dissertation, the related work section will focus specifically on the use of AR for healthcare providers and professionals.

2.2.1. Education and Training

Emergency Aid and Trauma Response Training

Over the past decade, the adoption of AR in emergency aid and trauma education has transformed from conceptual innovation to functional integration in both academic and clinical settings. This trajectory reflects growing confidence in AR's capacity to simulate high-risk scenarios, enhance procedural precision, and facilitate distributed learning environments for both professionals and caregivers. Several approaches have been investigated to improve the effectiveness of CPR training, ranging from traditional classroom-based instruction to advanced technological interventions, such as virtual reality simulations and AR applications [77, 119, 129, 143, 222, 279].

Initial research between 2015 and 2021 focused primarily on prototyping and evaluating the feasibility of AR platforms in emergency care training. One of the earliest efforts, Demir et al. [58], presented an AR-based mass casualty incident (MCI) simulation platform involving trauma nurses and emergency physicians, which helped users rehearse roles and response procedures in large-scale emergencies.

Mather et al. [170] tested AR for remote procedural guidance in rural trauma cases such as roadside emergencies and emergency childbirth. Their field trials demonstrated the value of AR in extending expert support to under-resourced areas through wearable smart devices.

A scoping review by Munzer et al. [187] systematically explored AR's potential in emergency medicine, identifying applications in CPR, trauma simulations, and emergency triage. The study highlighted AR's potential to reduce cognitive load, enhance procedural memory, and improve spatial awareness in high-stress clinical settings [187].

Simultaneously, Balian et al. [16] demonstrated that AR could improve cardiopulmonary resuscitation (CPR) training outcomes by providing visual prompts and correcting form in real time [16]. This was extended by Follmann et al. [85], who employed smart glasses in MCI triage simulations, allowing telemedical experts to supervise first responders remotely.

Leary *et al.* [149] focused on the limitations of Balian *et al.* work [17] and compared the use of CPRReality training with a standard audio-visual feedback manikin in terms of improvement in CPR quality. The findings of their study indicated that there was no statistically significant difference observed between the two groups. This implies that further, more extensive studies should be conducted to explore whether AR CPR training has the potential to enhance overall CPR quality for both new and re-certifying healthcare providers.

In 2021, multiple studies marked a turning point toward systematized AR

use. Ito et al. [120] integrated AR and extended reality (XR) into trauma care simulations within hybrid emergency rooms, enhancing multi-disciplinary coordination and visual injury assessment. Follmann et al. [86] developed an AR-based system to guide real-time clinical decision-making, especially helpful for inexperienced staff during triage scenarios. Around the same time, Schild et al. [230] launched "ViTAWiN," a mixed-reality trauma simulation platform designed for multi-professional medical teams, highlighting improvements in communication, accuracy, and collaborative behavior.

Building on this foundation, research after 2021 focused on real-world deployments, scalability, and integration with emerging technologies such as digital twins, AI, and IoMT.

McMullen et al. [173] led a national-scale intervention in Uganda, where AR training modules were deployed to upskill health workers and volunteers in trauma response. This study demonstrated AR's adaptability in low-resource and post-conflict healthcare systems, contributing significantly to emergency preparedness across remote regions.

Harari et al. [104] introduced AR into operating room crisis protocols, enhancing team adherence to checklists during simulated critical incidents. The AR interface guided team behavior in real time, improving outcomes through structured information overlays.

Further innovations came from [240], who explored the fusion of AR with artificial intelligence (AI) to simulate disaster preparedness scenarios in trauma units. This integration enabled predictive simulation loops and faster emergency response adaptation.

Meanwhile, Esmaeeli et al. [75] proposed a development framework for AR content in trauma anatomy education, specifying pedagogical and technical criteria for platform design. Complementary work by Faisal et al. [78] explored digital twin integration, allowing emergency training environments to adapt in real-time to the learner's performance and patient profiles.

Ajitha et al. [4] showcased an AR-powered triage platform capable of real-time health assessment and visual prioritization of patients, useful in chaotic trauma situations. Bignami et al. [22] provided a critical lens on this growth, urging institutions to adopt technoethical frameworks to regulate data use, trainee exposure, and AI biases in immersive simulations.

Another emerging frontier in AR-based emergency education is the integration of multimodal sensor fusion, combining biosensors, motion detectors, and wearable devices to enhance contextual realism and learner feedback. Antoniou et al. [9] conducted one of the earliest cohort studies utilizing biosensor-driven affective analytics within AR medical games. Their system monitored real-time emotional and physiological responses (e.g., heart rate, galvanic skin response) to adapt the training environment based on learner stress and engagement levels. Similarly, Demir et al. [58] incorporated environmental sensors and situational awareness tracking in their AR platform for mass casualty incident simulations, enabling personalized emergency scenarios based on real-time data.

More recent innovations have pushed toward wearable sensor integration.

For instance, Kim et al. [138] reviewed the rise of smart helmets combining AR displays with impact sensors, inertial measurement units (IMUs), and GPS, tailored for use in community trauma detection and triage. In a related vein, Li et al. [158] implemented data gloves with embedded pressure and flexion sensors in mixed reality environments for occupational emergency simulations. These modalities help replicate tactile feedback and assess fine-motor control, a crucial need in surgical or airway trauma simulations.

Another significant study by Innocente et al. [115] mapped tools enabling sensor-enhanced human–machine interaction in healthcare, especially for training clinicians in injury detection and real-time procedural feedback using AR overlays and real-time detector synchronization. Notably, Taha et al. [250] proposed photonics-powered AR skin electronics, which integrate optical sensors into training gear to visualize blood perfusion or injury severity during trauma simulations.

Despite these advancements, the field remains fragmented, and standardized frameworks for multisensor integration with AR interfaces are lacking. Future research must address interoperability issues, cost scalability, and ethical considerations around biometric data use in training environments.

Anatomy Visualization

AR has increasingly become a transformative tool in medical and health education, particularly in the teaching of human anatomy. Traditional methods, such as cadaveric dissection, while foundational, often present limitations in accessibility, repeatability, and cost. AR addresses these challenges by offering an immersive, interactive, and non-destructive platform for anatomical exploration.

Romero-Reveron [224] emphasizes the pedagogical shift in anatomy education, noting that AR tools allow for dynamic visualization of complex anatomical structures and better retention of spatial relationships among body systems. Chatha [40] explores this transition from scalpel-based dissection to AR and AI-enhanced simulations, arguing for a blended approach where AR augments, not replaces, traditional dissection. This is especially crucial in an era where dissection resources are becoming scarce.

In an interdisciplinary exploration, Warren et al. [194] showcase how AR, in conjunction with VR, supports not just anatomical training but broader clinical skills by enhancing learner engagement and comprehension. Similarly, Putro and Biddinika [214] demonstrate the effectiveness of markerless AR in modeling 3D representations of the human heart, providing an intuitive and portable method for teaching cardiovascular anatomy.

From a technical perspective, Ganesan and Poonkuntran [92] detail how semantic segmentation and geometric calibration techniques improve the fidelity of AR-based anatomical models, ensuring higher accuracy in visual overlays. Kumar et al. [142] expand on this by discussing the integration of AR with medical imaging, such as CT and MRI, enabling layered visualizations that merge live patient data with static anatomical frameworks for more personalized education.

Taken together, these studies illustrate that AR is not merely a visualization aid but a pedagogical innovation that supports deeper learning through multimodal, real-time interactivity. The current trajectory suggests that AR will play an increasingly vital role in democratizing anatomy education, particularly in resource-constrained or remote training environments.

Surgical Training

The integration of AR into surgical education represents a pivotal advancement in the simulation-based training paradigm. By overlaying digital content on real-world environments, AR allows surgical trainees to visualize internal anatomy, practice techniques, and receive real-time feedback, all within a risk-free environment.

Sampaio et al. [227] developed an AR platform specifically for remote surgical training during health emergencies and natural disasters. Their system enabled interactive visualization of procedures and was designed for deployment in low-resource and crisis settings, highlighting AR's flexibility in decentralized learning models. This aligns with broader shifts toward immersive, location-independent medical education.

A key concern in adopting such technologies is the need for structured training frameworks. Puliatti et al. [41] emphasize that the use of advanced imaging and robotic interfaces in urological surgery demands not only access to AR platforms but also a specialized curriculum that fosters skill translation from simulation to the operating room. Similarly, Yang et al. [275] demonstrated that ophthalmic residents using extended reality (XR) simulators achieved enhanced procedural precision and confidence, particularly in microsurgical tasks.

Maqsood et al. [167] conducted a systematic review on the use of AR for telestration in surgery, where instructors annotate in real time over surgical views. Their findings suggest AR-based telestration can significantly enhance surgical mentorship and remote instruction, reducing time to competence and improving surgical communication.

From a broader perspective, Bhatti et al. [21] examined AR and VR in radiologic guidance and image overlay techniques, finding increased diagnostic clarity and improved surgical navigation accuracy, particularly in minimally invasive procedures. The implications extend beyond traditional surgery, reinforcing AR's role in enhancing both preoperative planning and intraoperative decision-making.

Finally, the role of the metaverse in medical training is emerging, as discussed by Burlacu et al. [35]. They propose a convergence of AR with immersive social environments where surgical trainees and mentors can collaborate globally in real-time, a development with profound implications for the democratization and scalability of surgical education.

Recent literature highlights a significant evolution in the technical implementation of AR for surgical training. One of the earliest large-scale reviews by Barsom et al. [19] categorizes AR systems into laparoscopic surgical training, mixed reality training, and fully immersive models, noting that high-fidelity

haptic feedback and optical marker tracking are essential components for accurate simulation. Similarly, Dickey et al. [61] introduced a urology training tool integrating fiducial markers and real-time 3D anatomical overlays, showing substantial improvements in trainee spatial understanding.

Williams et al. [270] conducted a systematic review of AR applications, noting that only a minority of studies used validated performance metrics, underscoring the need for robust evaluation frameworks. Meanwhile, Aydin et al. [249] focused on software pipelines, identifying Unity3D and ARToolKit as leading platforms for surgical simulation development. They also emphasized the need for latency optimization to prevent motion sickness in high-frequency simulations.

In open surgery, Fida et al. [80] examined various tracking modalities such as inertial measurement units (IMUs), infrared (IR) sensors, and SLAM-based depth mapping to increase the reliability of AR overlays. Lahanas et al. [147] compared AR systems against traditional box trainers, concluding that AR enhances psychomotor learning when paired with task-specific calibration.

Viglialoro et al. [264] present a hybrid laparoscopic simulator using AR to guide identification of anatomical landmarks during cholecystectomy. Their platform integrates ARCore for spatial tracking and uses a real-time segmentation algorithm to detect and isolate tissue structures dynamically. Zhang et al. [282] explored MR headsets in emergency surgical procedures, finding that gesture-based interaction improves trainee autonomy.

Moreover, Lungu et al. [162] and Vávra et al. [261] emphasize modularity and interoperability, especially with intraoperative navigation systems and robotic surgical platforms. These studies advocate for a layered architecture approach, where AR modules communicate with imaging systems via DICOM protocols and are optimized using GPU acceleration.

Collectively, these technical advancements demonstrate that AR in surgical education has matured from static model overlays to dynamic, responsive platforms that integrate real-time anatomical data, motion tracking, and remote instruction. However, consistent benchmarking, data privacy compliance, and cost-effectiveness remain key barriers to widespread implementation.

Overall, AR is reshaping surgical training by increasing access to high-fidelity simulation, enabling personalized learning, and supporting continuous mentorship, though its effectiveness depends heavily on curricular alignment, technology access, and learner engagement.

2.2.2. Clinical Support Systems

Surgical Domain

AR is increasingly recognized as a powerful tool for supportive and assistive applications in surgery, spanning the full perioperative continuum, especially in orthopedic, plastic, and neurosurgery [45, 68, 101]. Numerous studies have demonstrated the feasibility and potential benefits of integrating AR technology compared to traditional surgical methods. These advantages include increased surgical precision and accuracy, improved anatomical awareness, re-

duced radiation exposure for both patients and surgical staff, shorter operative times, improved perceived workload for surgeons, and educational enhancements [2, 101].

In the context of preoperative planning, Bui et al. [34] reviewed the application of immersive AR platforms in spine surgery and emphasized their value for spatial orientation, risk visualization, and anatomical rehearsals. Their findings suggest AR tools are particularly effective for planning minimally invasive procedures, where anatomical variability can be a significant challenge. Similarly, Louis et al. [161] presented a case series demonstrating the integration of a synchronized AR platform with neuroimaging tools for operative neurosurgery. Their system allowed real-time visualization of tumors and vascular landmarks directly on patient-specific anatomy before incision.

Surgical rehearsal is another area where AR has shown significant promise. Bui et al. [34] and Sun et al. [248] report that procedural walk-throughs using AR systems help improve muscle memory, procedural sequencing, and confidence among residents. In hip surgery, for example, AR allowed trainees to rehearse acetabular exposure and implant placement, which are traditionally difficult to visualize in cadaver-based training. These rehearsal environments also facilitated objective performance assessment by tracking motion metrics and precision.

In terms of intraoperative guidance, Shen et al. [234] detail recent advances in AR-assisted visual overlays that dynamically adapt during surgery. Their work highlights the integration of optical sensors and real-time data to update the AR display, improving navigation through tissues and reducing the reliance on verbal or textual instructions. In orthopedic surgery, Jud et al. [128] conducted a systematic review and found AR-enhanced navigation improved implant alignment and reduced fluoroscopic exposure time across various procedures.

Unlike the aforementioned surgical specialties, implementing an AR assistance and navigation system for open abdominal surgery has encountered challenges, primarily due to organ mobility and deformations [81]. These difficulties arise from factors such as respiratory motion, positional changes resulting from laparotomy[1], or mobilization and deformations during intraoperative dissection [225, 252]. Nonetheless, these displacements and deformations are less pronounced in pancreatic surgery due to its special anatomical location. Despite this, there is still a limited number of studies in the literature that explore the design process and effectiveness of AR-based assistance and navigation systems during pancreatic surgery [168, 195, 200, 252].

Tang et al. [252] introduced an AR navigation system using a smartphone display during Pancreaticoduodenectomy[2] procedures in three patients. Their

[1] Laparotomy is a surgical procedure that involves making a large incision through the abdominal wall to gain access to the abdominal cavity.

[2] Pancreaticoduodenectomy, also known as Whipple procedure, is a complex surgical operation typically used to treat pancreatic cancer or other conditions affecting the pancreas. During the procedure, the surgeon removes the head of the pancreas, the duodenum (the first part of the small intestine), part of the bile duct, the gallbladder, and sometimes part of the stomach.

AR system effectively assisted in navigating and identifying the location of the superior mesenteric vein and the area of tumor invasion, with the deviation between the tumor invasion area and its representation in the 3D model estimated to range from 2 to 8 mm. Similarly, Marzano et al. [168] and Okamoto et al. [195] utilized AR navigation systems where reconstructed 3D models were overlaid onto real organs via a monitor display. These studies demonstrated that AR facilitated the precise and safe identification of the pancreas, tumor location, and critical vascular structures, with registration errors ranging from 5 to 12 mm [195]. Onda et al. [200] conducted the first comparative analysis between AR-based systems using 2D displays and conventional pancreatic resection procedures, reporting no complications associated with the AR system during or after the surgeries. Their study showed that the AR navigation system provided accurate anatomical information, allowing for the swift identification and ligation of important vessels.

In conclusion, while there are studies showing promising results for using AR to visualize patient data preoperatively [163, 221] and to assist with surgical planning [209, 220, 221, 239], the use of AR during surgery has received less attention due to the significant challenges involved. Additionally, most prior research in medical AR has primarily focused on assessing the efficacy of these systems, with an emphasis on system accuracy, often tested in controlled lab environments [53, 81, 82, 191] or in high-fidelity scenarios involving phantom patient cases [109, 218]. While these studies are essential for testing and evaluating new approaches prior to clinical use with actual patients, most systems' performance and usability have not yet been thoroughly evaluated in real-world surgical settings with actual patients. This highlights a critical gap in the validation process, as the practical application of AR systems in clinical environments remains largely untested, leaving questions about their functionality, reliability, and user acceptance in real surgical contexts. Although accuracy is undoubtedly crucial for advancing AR applications in this critical field, there has been a notable lack of emphasis on user-centric design and the testing of these systems in real-world settings. The feasibility of these technologies, particularly from the perspective of the surgeons who use them, remains underexplored, signaling an essential area for further investigation and development.

Other Domains

Beyond surgical settings, AR is playing a growing role in enhancing clinical workflows, particularly in areas like patient data visualization, real-time remote collaboration, and medical record interaction. Birlo et al. [24] performed a systematic review on optical see-through head-mounted displays, finding AR significantly enhanced real-time access to clinical data and enabled overlaying of digital records within the physician's field of view. This improved information retention and procedural safety in diagnostic environments.

Kim et al. [136] developed a collaborative AR-based imaging system tailored for group consultations in congenital heart disease, where multiple clinicians could simultaneously explore 3D heart models and review case files in

a synchronized virtual environment. Similarly, Negrão [190] emphasized the potential of mixed reality interfaces for remote multi-user surgical planning and teleconsultation, noting key design challenges around latency and fidelity in shared augmented spaces.

In neonatal and orthopedic care, Tueni and Amirouche [257] illustrated the application of AR in patient monitoring and procedural rehearsals. They proposed an interactive framework where 3D holographic patient data can be dynamically updated during multidisciplinary rounds. Rodrigues et al. [223] further highlighted AR's role in neonatal care, using mid-air gesture controls to interact with vitals, scan data, and alerts, minimizing physical contact and improving sterile workflows.

Chapter 3

Designing AR for Human Performance: Prerequisites of Viable AR

Contents

As AR technologies continue to advance, their integration into critical domains such as medicine is increasingly anticipated, yet not without substantial challenges. While AR holds the potential to enhance precision, reduce cognitive load, and support complex decision-making, its successful deployment hinges on a nuanced understanding of the technological and human factors that define high-stakes medical environments. This chapter explores the essential prerequisites for implementing AR in such contexts, focusing on two pivotal considerations: the choice of head-mounted display (HMD) technology, particularly the trade-offs between optical see-through (OST) and video see-through (VST) devices, and the design of interaction modalities that align with clinical needs and constraints.

The content of this Chapter covers two publications: .

1. Javaheri, Hamraz, et al. "When AR Hinders Performance: The Hidden Costs of Video-See-Through Displays." Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology. 2025.

2. Javaheri, Hamraz, Vitor Fortes Rey, Paul Lukowicz, Gregor Alexander Stavrou, Jakob Karolus, Omid Ghamarnejad, "Assessing the Feasibility of Using Apple Vision Pro While Performing Medical Precision Tasks: A Controlled User Study", JMIR XR AND SPATIAL COMPUTING (JMXR), 2025;2:e73574; doi: 10.2196/73574

3.1. Hardware Shapes Performance

Mixed reality (MR) has been a transformative technology for several years, revolutionizing various industries and applications. As part of the broader spectrum of immersive technologies [182], MR bridges AR, which overlays digital content onto the real world, and virtual reality (VR), which provides fully immersive digital environments. With recent advancements in wearable technology and head-mounted displays (HMDs), MR has expanded into a wide range of daily activities and professional domains [55]. A significant development in this field is the rise of extended reality (XR) devices, which integrate both AR and VR capabilities, enabling seamless transitions between immersive and real-world experiences. This new generation of devices, such as the Apple Vision Pro (AVP) [10], has sparked considerable interest and is believed to be the future of HMDs in medical domain, offering users immersive XR experiences through video-see-through (VST) technology [72, 73, 169, 198, 201, 228, 233, 255, 267].

Despite the promising potential of VST XR devices such as the AVP, a significant gap remains in understanding their usability in precision-critical domains, where tasks require high levels of real-world accuracy and dexterity. While studies have examined XR applications across various settings and domains, the choice of device and technology for specific task groups is often driven by the latest market trends rather than an informed assessment of their feasibility for the intended use case. This decision becomes particularly crucial for delicate tasks requiring fine motor control and precision, where the device itself may impact task performance, regardless of the XR application or method employed. The lack of empirical insights into the suitability of different XR technologies and devices poses a major barrier to the widespread adoption of XR in applications demanding precision and dexterity.

This section looks into the feasibility of the latest head-mounted displays with optical see-through and video see-through designs and presents their impact on user task performance and dexterity.

3.1.1. Methodology

Using a mixed-methods research approach, combining both quantitative and qualitative data collection techniques, two user studies were conducted, each experiment focusing on a different task, including a pegboard dexterity task and a medical-focused task of wound suturing. The aim of the studies was to evaluate the feasibility of the latest XR devices, Apple Vision Pro [11], and Microsoft HoloLens 2 for performing precision tasks and to analyze each device's impact on users across different measures. In each experiment, the same three study conditions were designed, corresponding to the three scenarios of the studies: Baseline, MHL2, and AVP. The Baseline condition, which involved no HMD, served as a reference point for evaluating the effects of the other conditions.

A within-subject design was employed, meaning each participant experienced all three conditions. To address potential order bias, the sequence of conditions was counterbalanced across participants.

User Study 1: Purdue Pegboard Dexterity Task

The initial experiment focused on user dexterity using the Purdue Pegboard Test (PPT), a psychomotor test of unimanual dexterity and bimanual coordination [241], focusing on two different abilities in user gross movements of arms, hands, and fingers, and fine motor extremity. The test is a manipulative dexterity assessment originally developed to help select candidates for industrial roles that demand precise hand skills, such as assembly, packaging, and machine operation. It evaluates performance separately for the right hand, left hand, and both hands simultaneously. The test measures two distinct types of dexterity: one involving larger movements of the hands, fingers, and arms, and the other focusing on fine fingertip coordination essential for tasks like small assembly work [254]. The PPT was chosen due to its extensive use in the literature as an assessment tool beyond its original intent, such as in rehabilitation [59], neuropsychological assessment [219], occupational medicine [159], and as an indicator of user performance and accuracy, helping to investigate the impacts of different interaction modalities [265] and other factors in immersive environments [1].

Study Task: The PPT consists of five subtests that evaluate fine motor skills and hand coordination. We conducted the following subtests in a standardized manner:

- **Right Hand (30 seconds):** Participants used only their right hand to place as many pins as possible into the right-hand row within 30 seconds.
- **Left Hand (30 seconds):** Participants used only their left hand to place as many pins as possible into the left-hand row within 30 seconds.
- **Both Hands (30 seconds):** Participants used both hands simultaneously to place as many pairs of pins as possible into both rows within 30 seconds.

- **Summed Score (Right Hand + Left Hand + Both Hands):** We calculated the sum of the scores from the first three subtests.
- **Assembly (60 seconds):** Participants used both hands to assemble a sequence of pins, washers, and collars within 60 seconds.

Participants completed three trials per subtest, with their average score recorded for analysis. This approach minimized practice effects and enhanced reliability, as three-trial administrations have been shown to yield higher reliability (0.76 – 0.89) after a one-week retest interval (0.81 – 0.89) [33] and maintain a reliability of 0.76 for peg-placing trials over six months [67].

Measurements: For each PPT trial, we recorded scores as follows:

- Right Hand, Left Hand, and Both Hands: The number of pins correctly placed within the allotted 30-second duration per subtest.
- Summed Score: The total calculated from the three individual subtests, including Right Hand, Left Hand, and Both Hands.
- Assembly: The number of correctly assembled pin-washer-collar units completed within 60 seconds.

To assess key aspects of the user experience, we administered online questionnaires. We collected data on demographics and participants' affinity for technology interaction (Ati) [88]. Additionally, cognitive workload was measured using the NASA-Task Load Index (TLX) [105], VR-induced sickness was evaluated with the VR Sickness Questionnaire (VRSQ) [139], and the sense of presence in the virtual environment was gauged using the Presence Questionnaire (PQ) [272]. Furthermore, the UMUX-Lite questionnaire [156] was employed for a quick and reliable assessment of new technologies [28], with system usability scores predicted using a regression equation [156] based on the two UMUX-Lite items.

Building on previous work using different sensing modalities for activity recognition [46, 245, 246], we used Apple Watches to record participants' movements for the objective assessment of study conditions. Our goal was to evaluate the impact of each device on unimanual movement during the Right Hand and Left Hand subtests, as well as bimanual movement during the Both Hands and Assembly subtests. IMU data were collected at a sampling rate of 100 Hz. Sample signals are provided in the supplementary file.

Furthermore, we conducted semi-structured interviews with participants. The interview questions covered topics including reflection on the experience, comfort, self-assessment of performance, advantages and disadvantages, and potential use of each device.

Study Procedure: The study began with an introductory phase, during which informed consent was obtained from participants, and an overview of the study tasks was provided. Following consent, the entire session was recorded

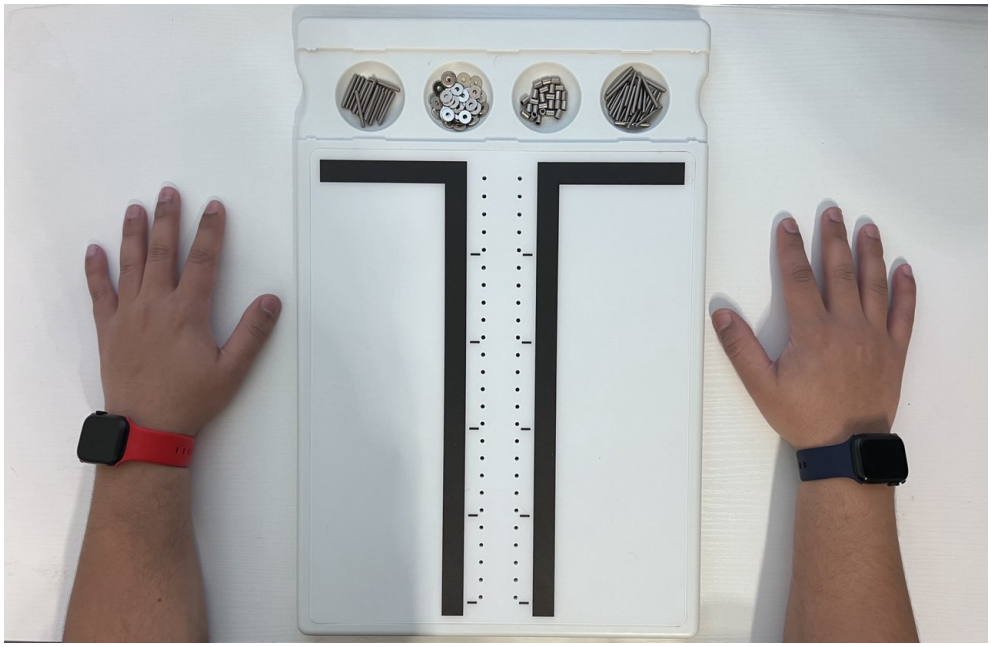


Figure 3.1.: The study setup, including the PPT board and two Apple watches placed on the participant's wrists.

using two cameras: one capturing an egocentric perspective and the other providing an exocentric, front-facing view. Additionally, an Apple Watch was connected to each of their wrists (Figure 3.1). The position of Apple watches remained the same throughout the entire experiment session.

Participants were then assigned to one of six possible sequences for completing the three conditions (Baseline, MHL2, and AVP) in a counterbalanced order. Before beginning the tasks, participants filled out two questionnaires on demographics and Ati [88].

After completion of the preparation phase and order assignment, the experimental task started. During this stage, participants were asked to perform the same four subtests (Right, Left, Both, and Assembly) under each of the three conditions (Baseline, MHL2, and AVP). Before each subtest, the procedure was explained and demonstrated to ensure participants understood the task. The participants were then asked to perform the subtest by placing a few pins for practice.

Before performing the task with the AVP and MHL2, calibration procedures were carried out. For the MHL2, only eye calibration was conducted, while for the AVP, both eye and hand calibration were performed to account for potential issues such as lens misalignment or visual discrepancies.

For each trial, the participants' scores were recorded for each subtest upon their completion. After completion of three trials of all subtests, online questionnaires were administered, asking participants to reflect on their experience regarding the completed study's condition. The order of the questions was randomized. After completing the Baseline condition (without an HMD), participants only filled out the NASA-TLX. For MHL2 and AVP conditions,

Table 3.1.: The study participants' demographics. M: Male, F: Female, NB: Non-Binary, R: Right-handed, L: Left-handed, A: Ambidextrous.

Characteristics	Value
Gender	14M (70%), 6F(30%), 0NB
Age	27.55 ± 5.31
Handeddness	18R (90%), 1L (5%), A(1%)
Prior usage of OST HMDs	1.75 ± 0.72 (1-5 Likert scale)
Prior usage of VST HMDs	1.85 ± 0.99 (1-5 Likert scale)
Affinity for technology interaction	3.87 ± 0.73 (1-6 Likert scale)

participants completed the NASA-TLX, VRSQ, PQ, and UMUX-Lite.

At the conclusion of each session, a researcher conducted a brief semi-structured interview with participants, inviting them to reflect on their experiences with each device.

Participant Recruitment: The study included only participants without vision disorders or those with minor refractive errors who could complete the tasks without eyeglasses. While MHL2 allowed for eyeglass use, AVP's design made this impractical. Although AVP could be adapted with corrective lenses, customizing them for each participant was not feasible. To ensure consistency and minimize bias, participants who had minor refractive error and typically wore glasses were required to complete all tasks without them across all study conditions.

A total of 20 participants volunteered for this study. Questionnaire outcomes were obtained from all 20 participants, and their demographic characteristics are summarized in Table 3.1. However, only IMU data from 13 participants were eligible for inclusion in the sensor data analysis. The remaining 7 participants were excluded due to data corruption. Exclusion criteria included any missing data windows for any subtests within any of the study groups.

Data Analysis: The sample size was determined using G*Power [69], a power analysis tool. Anticipating significant effects of different HMDs on user performance and cognitive load during study tasks, a large Cohen's effect size of 0.40 was assumed [48]. With a power of 0.95 ($1–\beta$) and a significance level of α = 0.05, the required sample size was calculated to be 18 participants. To account for a potential 10% dropout rate, 20 participants were recruited for the study.

Statistical analysis was performed using the R Project for Statistical Computing [215]. Continuous data were presented as means ± standard deviations, while categorical data were summarized as frequencies and proportions. For all numerical analysis (including dexterity scores from PPT subtests and questionnaire outcomes), the effects of the three conditions (Baseline, MHL2, and AVP) were analyzed using a within-subject repeated-measures ANOVA. When

the sphericity assumption was violated ($\epsilon < 0.75$), the Greenhouse-Geisser correction was applied. To address multiple comparisons, P-values were adjusted using the Bonferroni correction, with a two-sided P-value of < 0.05 considered statistically significant.

In addition to our statistical analysis of dexterity scores and questionnaires, we also collected IMU data from both hands for statistical analysis. To measure the overall rate of movements during PPT subtests using collected accelerometer data from hands, we computed the root mean square (RMS), that is,

$$\sqrt{\frac{1}{N \cdot C} \sum_{i}^{N} \sum_{c}^{C} a_{i,c}^2}$$

where $a_{i,c}$ is the acceleration value at time i for channel c. This metric was used based on previous work demonstrating the effectiveness of RMS on analysis of repetitive movements such as gait analysis [231]. We computed the RMS for each hand involved in a PPT subtest per trial. Since we used clapping as a synchronization gesture between the IMU and video data between sessions, we cropped the first and last 1.5 seconds so that the gesture would not influence the RMS computation. The generated RMS values were then used for statistical analysis to observe differences between study conditions. For the PPT subtests Right Hand and Left Hand, we only used data from the respective hand for analysis. For the Both Hands and Assembly subtests, data from both hands were used. We fitted linear mixed-effect models for the RMS for each PPT test using condition as a fixed effect and a by-subject random effect. All interviews conducted in this study were transcribed verbatim. A pragmatic approach to qualitative analysis was employed, following the recommendations of Blandford et al. [26]. Initially, two researchers independently analyzed 25% of the interview data. Through iterative discussions, a preliminary coding framework was established. The remaining interview data were then divided equally between the two researchers for coding. In a final discussion, the coding framework was further refined, culminating in the identification of key themes.

User Study 2: Wound Suturing Task

Study Task: To assess the feasibility of different HMDs and their impact on users while performing precision-dependent medical applications, we designed a controlled user study task that incorporates performing different suture techniques. The study included three types of sutures, each requiring progressively more complex techniques ranging from basic to advanced. The simple interrupted suture (SIS, Figure 3.2 (A)) was selected as the simplest task, while the vertical mattress suture (VMS, Figure 3.2(B)) and continuous subcuticular suture (CSS, Figure 3.2 (C)) were chosen for their complexity. These techniques rely on correct depth perception, as they involve inserting a suturing needle into a specific layer of the skin [148], making them ideal for evaluating users' ability to perceive depth in a simulated environment.

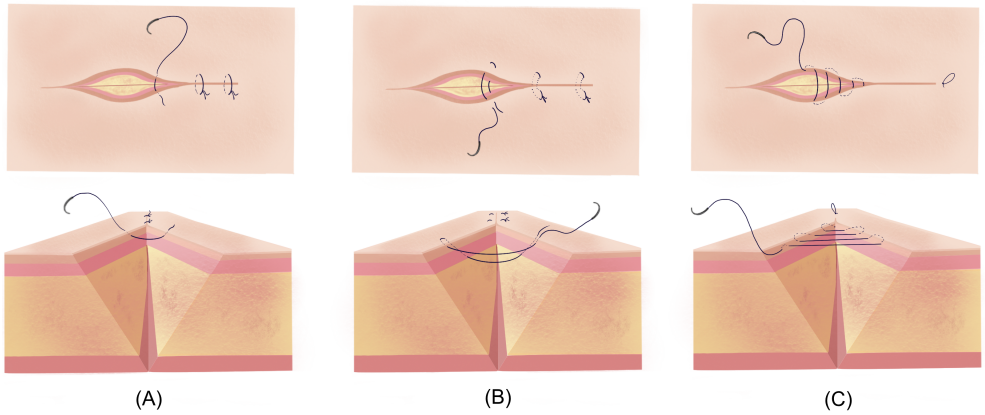


Figure 3.2.: Study 2: The illustration of three suture types included in the study task: (A) Simple interrupted suture, (B) vertical mattress suture, and (C) continuous subcuticular suture.

A suture training kit [174] was used as the base for performing the sutures. All participants were asked to complete three SIS, three VMS, and close a 5 cm long wound on the suturing kit using a CSS for each intervention. All sutures were performed using 3-0 polypropylene [76] and the same clinical surgical instruments, including needle holder, tweezers, and scissors (Figure 3.3). To minimize potential bias in TCTs due to the length of the suture material, each suture type was performed with a new suture material.

Measurements: The entire experiment session was recorded using two cameras, one egocentric and one exocentric front-facing views. After completing the task for each intervention, photographs of the participants' performance using the suturing kit were captured for subsequent evaluation. Additionally, online questionnaires were administered to evaluate key factors related to the user experience. Cognitive workload was assessed using NASA-TLX [105], VR-induced sickness was measured with the VR sickness questionnaire (VRSQ) [139], and the sense of presence in the virtual environment was evaluated using the presence questionnaire (PQ) [272]. Furthermore, for a rapid and reliable assessment of new healthcare technologies [28] UMUX-Lite questionnaire was used. System usability score then was predicted using a regression equation based on the two UMUX-Lite items [156].

After all data recording sessions were completed, the task completion times (TCTs) were extracted from the recorded videos. A researcher, who was blinded to the study's aims, measured the TCT for each suture performed by every participant. To ensure an objective evaluation of the time spent solely on suturing, TCT was defined as the duration from the moment the needle was grasped by the needle holder until the knot was cut with scissors. Additionally, five surgeons (7.2 ± 1.7 years of surgical experience), who were also unaware of the study aims, evaluated suture performances based on anonymized photographs of performed sutures. All of the captured pho-

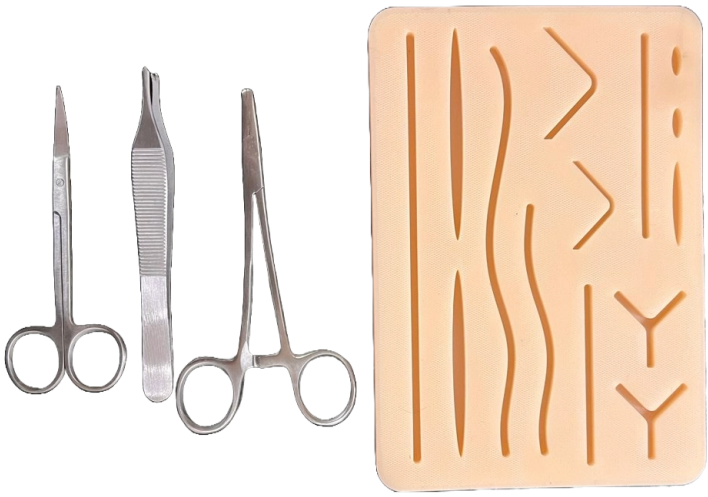


Figure 3.3.: Study 2: The suture training kit used during study tasks containing a silicone suture pad and instruments.

tographs of participants' performed sutures were presented in a random order to evaluator surgeons using a custom visualization tool. They rated the performance of each suture type separately on a scale of 0 to 100, considering factors such as the overall effectiveness of the suture, bite (length of the stitch across the wound), pitch (interval between stitches), and cosmetic appearance [148].

Study Procedure: The study began with an introduction phase involving obtaining informed consent and an introduction to the study tasks from all participants. Following consent, participants were randomly assigned to one of six possible orders for performing the three interventions (Baseline, MHL2, and AVP). Prior to task execution, participants completed two questionnaires on demographics and affinity for technology interaction [88]. Following the completion of the preparation and order assignment, the experiment task began. During this stage, participants were asked to perform the same study task, which involved working with three different suture types, for each intervention. Prior to performing the study task with AVP and MHL2, calibration procedures were conducted. For MHL2, only eye calibration was performed. For AVP, both eye and hand calibration were performed to address any potential issues arising from lens misalignment or visual discrepancies. After completing the task for the baseline condition (without an HMD), participants completed only the NASA-TLX. Following the MHL2 and AVP interventions,

participants filled out the NASA-TLX, VRSQ, PQ, and UMUX-Lite. Finally, at the end of each session, a researcher conducted a short semi-structured interview with the participants, asking them to reflect on their experience with each device. The interview questions encapsulated aspects including comfort, self-performance evaluation, pros and cons, and potential use cases.

Participant Recruitment: Participant Recruitment Recruitment was conducted through word-of-mouth and advertisements via mail. The experiment took place at Klinikum Saarbrücken, Germany. Participation was entirely voluntary, and no compensation was provided.

The inclusion criteria for participants required that they be healthcare professionals with prior experience in performing wound suturing on patients. Additionally, the study limited participation to individuals without vision disorders or those with minor refractive errors, who could complete the study tasks without eyeglasses. To minimize bias, participants with minor refractive errors who typically used eyeglasses were instructed to perform all tasks without glasses across all interventions.

In total, 20 healthcare professionals participated in this study. The demographic characteristics of the participants are detailed in Table 1.

Table 3.2.: The study 2 participants' demographics. M: Male, F: Female, NB: Non-Binary, R: Right-handed, L: Left-handed, A: Ambidextrous.

Characteristics	Value
Gender	14 M (70%), 6 F (30%)
Age (years)	33.65 ± 7.60
Occupation	Surgeon: 15 (75%) Physician assistant: 1 (5%) Medical intern: 4 (20%)
Participants with minor refractive errors	6 (30%)
Clinical experience (years)	7.8 ± 6.45
Prior usage of OST HMDs (1–5 Likert scale)	1.8 ± 0.93
Prior usage of VST HMDs (1–5 Likert scale)	1.35 ± 0.63
Affinity for technology interaction (1–6 Likert scale)	3.83 ± 0.79

Data Analysis: Sample size calculation was performed using the power analysis tool G*Power [100]. Given the anticipated substantial impacts of different HMDs on users' performance and cognitive load during suturing tasks, a large Cohen's effect size of 0.40 was considered. With a power of $1–\beta$ = .95 and α = .05, the required sample size was calculated to be n = 18 participants. To account for 10% potential dropouts, we included a total of 20 participants in the study. Statistical analysis was conducted using the R project for statistical computing [216]. Continuous data were expressed as means ± standard deviations, while categorical data were reported as frequencies and proportions. The effects of the interventions (baseline, MHL2, and AVP) were analyzed within-subject using repeated-measures ANOVA. When the assumption of sphericity

was violated, the Greenhouse-Geisser ($\epsilon < .75$) correction was applied. To account for multiple comparisons within groups, P-values were adjusted using the Bonferroni correction. A two-sided P value of ¡ .05 was considered statistically significant for all analyses. Furthermore, the inter-rater agreement of the performance scores by five evaluator surgeons was confirmed using the $r_{wG(J)}$ agreement index [122]. Qualitative Analysis All interviews conducted in this project were transcribed verbatim. We adopted a pragmatic approach to qualitative analysis, as recommended by Blandford et al. [26]. Initially, two researchers independently analyzed 25% of the data. Based on iterative discussions, a preliminary coding framework was developed. The remaining interview data were then evenly distributed between the two researchers for coding. In a final discussion, the coding framework was further refined, leading to the development of the main themes.

3.1.2. Impacts of Device Selection on User Performance and Dexterity

The results of both studies demonstrated that the device used while performing precision required task has a direct impact on the user's performance and dexterity.

The ANOVA test on average dexterity scores across all subtests in the user study 1 using PPT, yielded a significant difference between study conditions ($F(2,38) = 3.65$, $p = 0.036$). The pairwise analysis on average score across all subtests showed a significant difference between AVP (23.4 ± 3.56) and Baseline (25.3 ± 3.60, $p = 0.028$); however, it yielded no significant difference between AVP (23.4 ± 3.56) and MHL2 (24.4 ± 3.83), and between MHL2 (24.4 ± 3.83) and Baseline (25.3 ± 3.60). Participants scored significantly lower on all subtests using AVP compared to Baseline; only the Left Hand subtest yielded a significant difference between MHL2 and Baseline, and no significant difference was observed between AVP and MHL2 for any subtests (Figure 3.4).

Furthermore, this significant decrease in dexterity scores observed under the AVP condition across all subtests was accompanied by a marked reduction in RMS acceleration values relative to Baseline in the single-handed subtests (Right Hand and Left Hand), where the most repetitive hand movements occurred.

For RMS, the resulting linear mixed-effect models show a significant effect for Right Hand ($F(2,102) = 3.30$, $p = 0.04$) and Left Hand ($F(2,102) = 9.23$, $p < 0.001$). Post-hoc pairwise comparisons revealed a significant effect for Right Hand between AVP (0.593 ± 0.008) and Baseline (0.596 ± 0.009, $p = 0.0393$) as well as for Left Hand between Baseline(0.592 ± 0.006) - MHL2 (0.590 ± 0.006, $p = 0.0245$), and AVP (0.589 ± 0.005)- Baseline ($p = 0.0001$). The pairwise comparisons on RMS across all subtests are given in Figure 3.5.

These results suggest that the device may have introduced perceptual or motor constraints that interfered with natural hand movement, leading to slower and potentially more deliberate actions. In the MHL2 condition, a non-significant decline in dexterity scores was observed, which was similarly mirrored by a reduction in RMS acceleration, especially for single-handed sub-

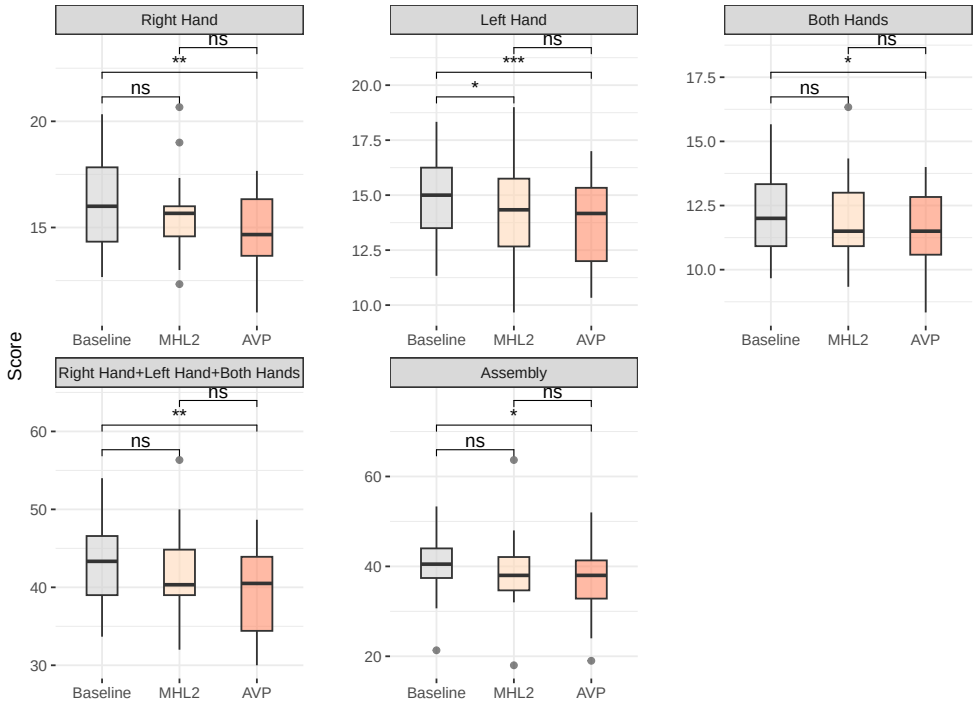


Figure 3.4.: Study 1: Pairwise result of participants' scores for each subtest across three study conditions (Baseline, MHL2, and AVP) with statistically significant differences marked. Significance levels: ns (not significant, $p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

tests. This deceleration in participants' movements was especially prominent in the Left Hand subtest (the non-dominant hand for 12 out of the 13 participants) for both devices. This indicates that the effect of the device on user performance may be exacerbated when using the non-dominant hand, possibly due to increased cognitive load.

A similar trend of reduction in participants' speed and dexterity while performing the suturing task was observed, which resulted in longer TCTs. The analysis revealed a significant difference in the TCT required to complete all suturing tasks across the interventions ($P < .001$). Participants took significantly longer to complete the entire task using AVP (570.0 ± 192.0 sec) compared to MHL2 (456.0 ± 120.0 sec, $P < .001$) and baseline (472.0 ± 143.0 sec, $P < .001$). Analysis performed on TCT for each individual suture type (Figure 3.6) revealed that there were no significant differences in the time required to perform SIS across the interventions (P = .05). However, a significant difference was observed for more complex tasks, VMS (P = .001), and CSS ($P < .001$). Participants required significantly more time to complete VMS using AVP (195.9 ± 79.6 sec) compared to MHL2 (161.2 ± 63.4 sec, $P < .001$) and baseline (166.8 ± 64.2 sec, P = .04). Similarly, for CSS, a significantly longer time was needed when using AVP (234.7 ± 83.6 sec) compared to MHL2 (177.0 ± 52.8 sec, P = .007) and baseline (182.9 ± 59.1 sec,

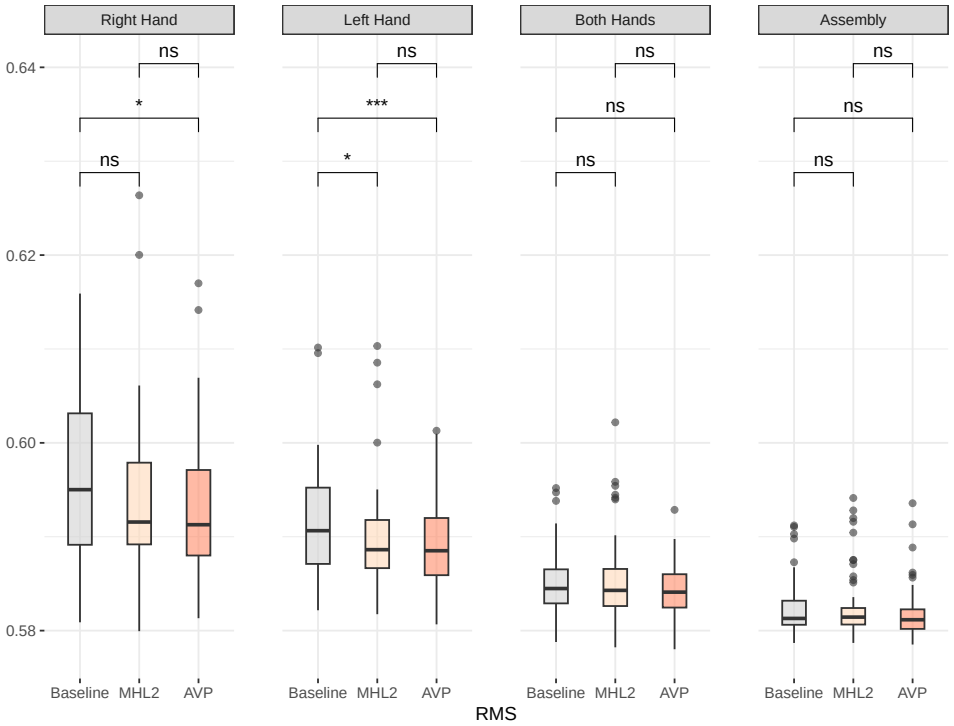


Figure 3.5.: Study 1: The statistical analysis on root mean square (RMS) values of IMU data across subtests with statistically significant differences marked. Significance levels: ns (not significant, $p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

P = .002).

Despite longer TCTs, the analysis on participants' average suture performance in user study 2 revealed no significant differences across interventions (P = .76). The five surgeons' evaluation scores of participants' suture performances showed high agreement, with an $r_(wG(J))$ value greater than 0.99 across the suture types. Despite the worsened performance with AVP for all suture types, no significant differences were observed between interventions for any suture types (Figure 3.7).

3.1.3. Impacts of Device Selection on User Cognitive Load

Both study results demonstrated that the OST or VST design of the used device has a significant impact on user cognitive load, regardless of the task being performed.

The study with PTT showed that the total raw (unweighted) TLX score among participants was significantly higher for AVP (46.9 ± 22.4) compared to MHL2 (32.8 ± 18.5, $p < 0.001$) and Baseline (30.9 ± 16.2, $p < 0.001$). The results on each individual factor of NASA-TLX are given in Figure 3.8.

Similarly, the analysis of study 2 also revealed a significant difference in the total raw NASA-TLX score across the interventions ($P < .001$). The total raw

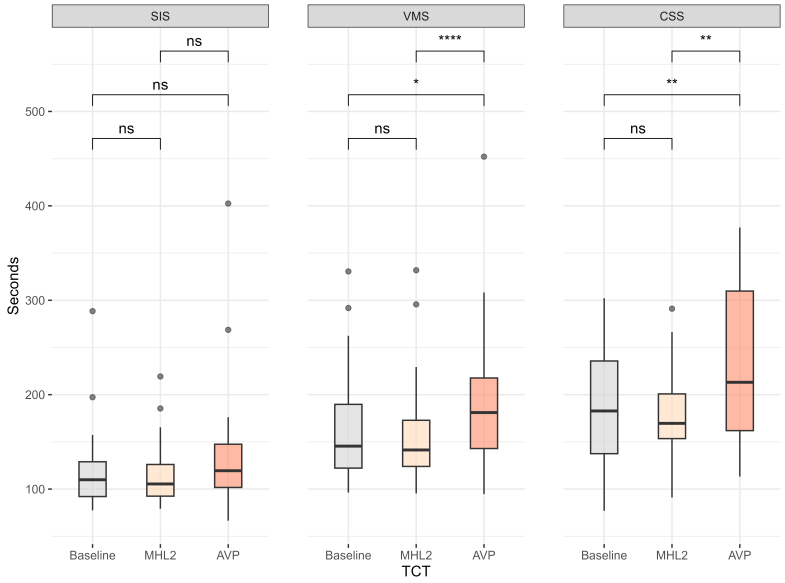


Figure 3.6.: Study 2: Participants' TCTs across different suture types. Statistically non-significant differences are marked with "ns", while significant differences are denoted as follows: $P < .05$ (*), $P < .01$ (**), and $P < .0001$ (****).

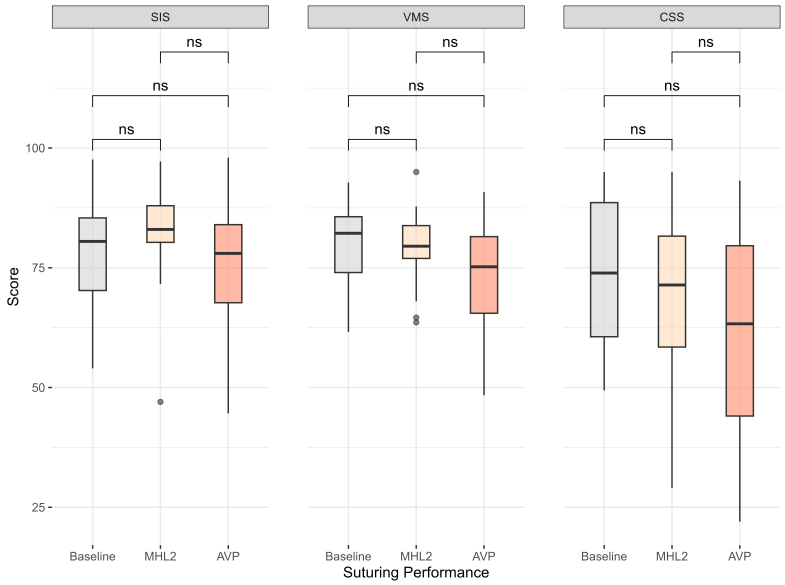


Figure 3.7.: Study 2: The participants' suturing performance rated by surgeons across different suture types. Statistically non-significant differences ($P > .05$) are marked with "ns".

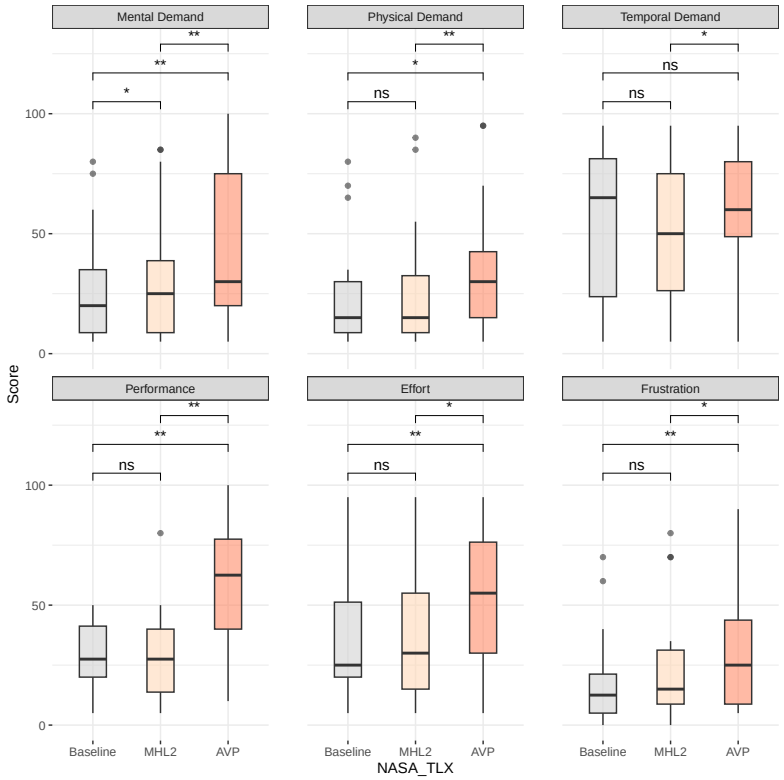


Figure 3.8.: Study 1: NASA-TLX results for each factor (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) across the three study conditions with statistically significant differences marked. Significance levels: ns (not significant, $p \geq 0.05$), * ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$).

NASA-TLX score among participants was significantly higher for AVP (43.9 ± 15.9) compared to MHL2 (21.5 ± 13.8, $P < .001$) and baseline (19.1 ± 15.1, $P < .001$). AVP scored significantly higher on mental demand, physical demand, effort, and frustration factors compared to MHL2 and baseline (Figure 3.9). There were no significant differences observed between MHL2 and baseline on any of the NASA-TLX factors.

3.1.4. User Comfort and Preferences

One of the main criteria in the integration of immersive technologies in the critical health-care domain is achieving user acceptance by ensuring practitioners can perform tasks confidently and without compromise. The device and technology should be used to enhance rather than hinder their capabilities. This encompasses the user emotions, physical comfort, and overall perception of the device's usability.

The analysis of the presence questionnaire in study 1 with PPT showed a significant difference in the presence score for MHL2 (93.2 ± 20.7) compared to AVP (84.5 ± 18.6, $p = 0.042$). The VRSQ results only yielded significantly

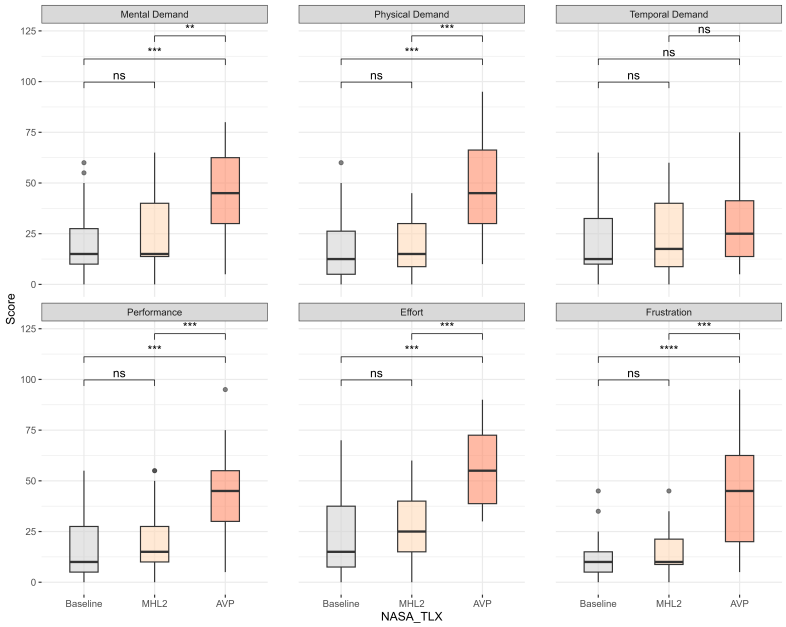


Figure 3.9.: Study 2: NASA-TLX results for each factor (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) across the three interventions. Statistically non-significant differences are marked with "ns", while significant differences are denoted as follows: $P < .01$ (**), $P < .001$ (***), and $P < .0001$ (****).

different results on the disorientation factor for AVP (61 ± 22.9) compared to MHL2 (49.3 ± 20.1, $p = 0.05$) and did not show any significant differences on the oculomotor factor for AVP (64.4 ± 30.3) compared to MHL2 (57.1 ± 22.0).

Similar outcomes to the study 1 were also achieved in the second study. The analysis of PQ demonstrated significantly higher presence score for MHL2 (115.0 ± 11.4) compared to AVP (93.7 ± 12.7, $P < .001$). This significant difference was observed in all factors of presence, including the realism ($P < .001$), possibility to act ($P < .001$), quality of interface ($P < .001$), possibility to examine ($P < .001$), and self-evaluation of performance (P = .03). The overall VRSQ score was also significantly higher for AVP (66.9 ± 19.8) compared to MHL2 (41.1 ± 9.32, $P < .001$). The participants rated significantly higher scores ($P < .001$) on both oculomotor (75.8 ± 22.9) and disorientation (58 ± 19.0) factors for AVP compared to MHL2 (42.9 ± 12.8, 39.3 ± 9.15).

Despite having different professional backgrounds and performed task in each study, both study outcomes also agreed in terms of a higher usability score for MHL2 with OST design compared to AVP. However, while the participants in the first study did not score devices significantly different in terms of usability (MHL2 = 61.9 ± 15.2, AVP = 56.8 ± 16.4), a significant difference was observed between usability scores given by healthcare professionals in the second study (MHL2 = 72.7 ± 8.54, AVP = 50.3 ± 14.4, $P < .001$). This difference in usability score might stem from having a different set of study

tasks and also the professional background of participants. While participants in the first task focused on the general usability of the device, as the task was not associated with any type of their professional background, the healthcare professionals in the second task rated the device with a focus on the usability for their very specific domain needs and criteria.

The thematic analysis of qualitative data in both study showed that, while both OST and VST has their own set of advantages, both study outcome agreed on that OST design remains a more suitable option for performing precision and dexterity required task specially for critical healthcare domains, while VST design might be more suitable for either entertainment purposes or rather in applications and domains where the user attention and focus is more on the virtual object rather than real-world interaction. This assumption was based on various reported factors, including comfort and physical strain, visual challenges, and self-evaluation of performance.

The participants in both studies repeatedly reported and discussed the comfort and physical strain as an important factor in their evaluation. The physical discomfort while using AVP is due to the device's unbalanced weight distribution, which concentrates pressure on the nasal and maxillary areas, leading to strain in the back neck muscles and headaches. Additionally, some participants in both studies experienced adverse effects such as eye strain, nausea, and dizziness, further diminishing their comfort and overall user experience

The sentiments such as:

> *"It (AVP) is not comfortable at all. It's a bit heavy on the head. And I felt like a bit of eye-strain. I couldn't focus much." (study 1, P8)*

> *"It (AVP) is very heavy, and after a while, you feel like your head would fall down if you don't consistently fight it. And I can assure you it was a relief to take it out." (study 2, P13, surgeon)*

> *"When you look straight, it is more convenient, but when you bend your head to look at the stitch pad, which I think would be the normal case when you operate around the table, then it is too uncomfortable, because the whole weight is in front and there is a constant contraction on your neck." (study 2, P9, surgeon)*

are indicators of a lack of user acceptance and reduced tendency to continue using the device due to device design, regardless of the application.

On the other hand, the positive comments such as:

> *"It (MHL2) didn't bother me, just like the glasses were there, that's it. Just kind of like wearing normal glasses ... Honestly, at some point, I kind of stopped noticing it." (study 1, P16)*

> *"With the HoloLens, after a few seconds, I didn't even notice. It felt like I was just doing the task. I was able to concentrate. It didn't really hinder or bother me." (study 1, P2)*

> *"We didn't use it here, but I guess with HoloLens you also have this option where you could push the visor up or down based on what you want to see, but with AVP you don't even have that option. You just have to take it off completely. And it is just not practical to use it during operation if you have to take it off every time." (study 2, P1, surgeon)*

demonstrated the user acceptance and tendency to use the device further in the case of the presence of a valid and useful application.

Beyond physical comfort, the visual challenges inherent in the design of VST AVP systems affected participants' perceptions of where this technology can be used. The limited field of view, difficulty focusing on fine details, visual distortions caused by head movement, and impaired depth perception were all cited as major barriers to using AVP for tasks that require precision and delicacy. This concern emerged consistently across both studies; although the task sets differed, participants encountered similar visual challenges that are believed to have undermined their performance.

> *"The AVP was harder to see through. If it's like a bigger target or like I don't need details, it's fine. But if I'm doing a task like an assembly task, I could barely see the little things." (study 1, P13)*

> *"It (AVP) was blurry because I tried to see some words in the Apple Watch. I couldn't see anything, and I was quite like, okay, is my eyesight that bad? Then, when I tried to do the same thing through the MHL2 display, I could see it clearly." (study 1, P18)*

> *"For the first two sutures, AVP was also ok, but for the last suture type, you have to really see where you put your needle in and out, which was simply impossible with AVP to make sure you are in the correct layer." (study 2, P11, physician assistant)*

> *"I used AVP before, but I only used it to watch videos, and initially I thought AVP would be better compared to HoloLens, but it was a complete catastrophe. I almost saw nothing. Yes, I did sutures from experience, but it was a complete guesswork, especially for the last suture type." (study 2, P1, surgeon)*

Interestingly, the increased blur caused by the camera's limited focus on very close objects in the VST design produced a counterintuitive effect. Rather than providing clarity when details were needed, the view grew less sharp up close. In both studies, participants noted that objects appeared sharper when held at a greater distance, but that distance was impractical for fine work, since they instinctively leaned in to improve their observation and precision.

> *"I could see things clearly farther than one meter from me. I could even read small letters, but when I looked at the stitching pad,*

> *which was closer to me, it became blurry. And got even blurrier when I was leaning closer to it to do the stitches. Which you usually should see better when you get closer, but it just made it worse." (study 2, P13, surgeon)*

Moreover, viewing the real world only through a video feed rather than directly was reported as another drawback of VST designs. Healthcare professionals felt less confident in carrying out accurate, real-world tasks, and they raised concerns about reduced social connections with both teammates and patients in the case of the usage of VST design rather than OST.

> *"With HoloLens I felt more secure and also more confident doing the task, because I think my peripheral vision was not affected that much, but in AVP, even though you still see, you have more restricted peripherals." (study 2, P2, surgeon)*

> *"I think when you talk to a patient or your colleagues around the surgery table, it feels just more natural to have eye contact even if it is through a glass like this (showing MHL2). It is better than having a big headset on your face where no one can see your eyes in there. It is just more assuring with a see-through glass than a completely closed one (AVP)." (study 2, P18, surgeon)*

3.1.5. Discussion

Our comparative evaluation of the Apple Vision Pro (AVP, a video-see-through or VST device) and Microsoft HoloLens 2 (MHL2, an optical-see-through or OST device) alongside a no-HMD baseline highlights the critical importance of device selection for precision tasks in healthcare. Across two user studies, we observed that AVP significantly increased cognitive load (NASA-TLX) and task completion times (TCT) compared to MHL2 and baseline, suggesting that the VST design imposes additional perceptual and attentional demands. Remarkably, despite these challenges, surgeon-rated performance scores showed no significant differences across conditions, yet participants' self-reported performance (captured in the Presence Questionnaire's self-evaluation subscale) declined markedly with AVP, reflecting reduced confidence and corroborating qualitative reports of visual difficulty and discomfort (e.g., eye strain, neck fatigue).

This divergence between objective and subjective performance echoes findings by Olexa et al. [198], who demonstrated AVP's utility for non-time-sensitive surgical planning, and suggests that AVP remains feasible where real-world precision and rapid decision-making are less critical. For example, AVP could support virtual simulations in medical education or telemedicine scenarios where interaction with virtual content is prioritized over exacting motor control. Conversely, our data align with Jud et al. [128] and Shen et al. [234] in underscoring OST advantages, MHL2 maintained low cognitive workload, preserved peripheral awareness, and yielded TCTs and dexterity

scores statistically indistinguishable from baseline, affirming its suitability for intraoperative guidance and real-time surgical navigation.

Comparing our findings to other works highlights device-context interactions. Unlike Olexa et al. [198], who reported minimal eye strain under AVP, our participants' VR-sickness scores revealed heightened oculomotor strain and disorientation, likely due to their focus on real-world visual tasks rather than purely virtual engagement. This discrepancy underscores the influence of user attention: VST HMDs may excel when the primary task involves virtual object interaction (e.g., anatomy visualization [267]), but struggle when precise real-world perception is required. Furthermore, our RMS and dexterity analyses mirror trends in studies of mixed-reality ergonomics [99, 125], confirming that non-dominant hand performance is especially sensitive to device-imposed motor constraints.

These insights carry several practical implications for XR design in medicine. First, OST designs with transparent optics and balanced ergonomics should be prioritized for precision-critical, time-sensitive procedures, such as suturing, open surgery guidance, and robotic-assisted operations, where maintaining a clear view of the physical environment and team communication is essential. Second, VST devices should be reserved for contexts where full immersion or remote visualization supersedes direct motor accuracy, including preoperative planning [198], telehealth consultations, and VR-based training modules. Third, future hardware iterations must address current VST limitations by improving camera focus and depth-of-field for near-field work, enhancing weight distribution, and refining display resolution to minimize blur and expand usable field of view.

As XR technology converges toward hybrid VR/MR headsets in commercial products (e.g., Apple's latest XR models [11], Meta's Quest Pro [177]), these findings provide a framework for aligning device capabilities with clinical demands. Ultimately, tailoring XR hardware and interaction paradigms to the specific cognitive and motor requirements of medical applications will be paramount in translating lab-based innovations into safe, effective, and widely adopted tools in healthcare practice.

3.2. Interaction Under Constraints

Wearable AR systems have evolved to support a wide range of interaction modalities, encompassing both input and output (or feedback) mechanisms, all aimed at enhancing user experience across various applications. In critical medical domains, where system usability can directly impact outcomes, the choice of interaction modalities becomes one of the most fundamental prerequisites for successful AR integration, regardless of the specific application. This section aims to provide an overview of different interaction modalities that could be incorporated into an AR system for the medical domain based on the needs and limitations of the domain and the aimed system functionality.

3.2.1. 5W1H Framework

Designing systems that are intelligent, adaptive, and user-aware requires a deep understanding of the user, their context, and the interaction itself. The **5W1H approach**—*Who, What, When, Where, Why*, and *How*—offers a structured lens through which designers can explore and define these critical dimensions early in the design process.

Applying 5W1H helps shift the focus from technology alone to the *situated experience* of the user. Each question guides reflection on different aspects of the interaction, ensuring the resulting system is contextually sensitive and responsive to real-world needs.

- **Who** is using the system?
  Understanding user identity, cognitive state, goals, skills, and preferences is essential for designing personalized and accessible interactions. Systems should adapt to different user profiles and support inclusive experiences.

- **What** is the user trying to do?
  Defining the tasks and intentions behind interactions helps shape functionality that aligns with user goals. It encourages task-oriented design that reduces friction and cognitive load.

- **When** does the interaction occur?
  Timing plays a key role in user experience. Recognizing temporal patterns (e.g., time of day, urgency, or activity phase) enables systems to provide timely and relevant support or interventions.

- **Where** is the user located?
  Whether the user is in a physical environment (e.g., at home, on the go, in a surgery room) or a digital context (e.g., mobile, desktop, AR/VR), understanding this spatial dimension allows systems to tailor interfaces and interaction methods accordingly.

- **Why** is the interaction happening?
  Going beyond surface-level behavior, asking *why* uncovers user motivations and values. This drives meaningful interaction design that supports deeper engagement and goal fulfillment.

- **How** does the interaction unfold?
  This encompasses the interaction modality (e.g., touch, voice, gesture), interface design, and system feedback. It's about designing intuitive, usable pathways that align with human capabilities and expectations.

When these six questions are considered together, they guide the creation of *intelligent systems* that are not only functional but also *aware of the user's cognitive state, environmental context, personal profile*, and the *most appropriate interaction modality*. This results in interactions that feel natural, adaptive, and supportive, the hallmarks of human-centered design.

3.2.2. Designing Input: Aligning Control with Context

Depending on user circumstances and environmental conditions, systems in the medical domain can often be controlled either by a third person or directly by the user. While this decision is often influenced by physical constraints or limitations of the user's setup, it can also be intentional, depending on the system's intended use. For instance, in environments with strict sterilization requirements, such as in surgical settings, displays and monitors are often controlled by a third person (e.g., a nurse) following the surgeon's request. However, as the complexity and the need for detailed interaction with the system increase, direct user control is often preferred. This approach saves time and bypasses communication challenges. For example, surgeons typically interact directly with imaging devices like ultrasound in the operating room, while nurses focus on ensuring that sterilization protocols are met.

In contrast, in settings such as education, remote control of the system is not due to environmental limitations but is instead an intentional design choice to deliver controlled output to the user. In a controlled educational context, trainers typically operate the system remotely, while trainees remain primarily as observers with limited control. For example, in nursing education, trainers use control panels to simulate patient scenarios by modifying system outputs, such as changing a patient's vital signs. Nursing students then respond to the simulated conditions based solely on what is presented to them, with minimal direct interaction with the system.

While remote system control often faces challenges related to connectivity and real-time responsiveness, it typically allows for a broader range of usable input modalities compared to direct user control. This is because when the user is the sole operator, the choice of input modalities is heavily influenced by their physical setup and environment. In contrast, remote control interfaces can be situated in settings with fewer constraints, enabling the use of more diverse and flexible interaction methods.

For wearable AR systems, direct user input can be facilitated through external devices such as controllers, clickers, or buttons (e.g., keyboards), and foot pedals, or through body-based inputs involving the full body, head, hands, eyes, feet, voice, or brain, each tracked using appropriate sensing technologies. The optimal modality can vary significantly depending on the application domain, environmental constraints, user capabilities, and task-specific demands. To support design decisions, the following questions serve as a practical checklist to evaluate which interaction modalities are most suitable for a given use case. Further details on the pros and cons of each input modality are also given in Table 3.3.

Questions to Guide Input Modality Selection

1. **Is the user expected to have full use of both hands during the task?**
   Rationale: Many AR applications involve simultaneous physical

tasks (e.g., assembly, surgery). If hands are occupied or gloved, hand-based inputs become less feasible.

2. **Is the domain noisy or is voice interaction impractical?**
   Rationale: Environments with high ambient noise or confidentiality concerns hinder reliable voice recognition and social acceptability.

3. **Are the users visually focused on a primary task (e.g. surgery)?**
   Rationale: Visual distractions can interfere with critical attention-demanding tasks.

4. **Is the domain sensitive to privacy or social context (e.g., public spaces, therapy sessions)?**
   Rationale: Modalities like voice and large gestures may be socially intrusive or raise privacy concerns.

5. **Are the users seated or stationary, or will they be moving (walking, etc.)?**
   Rationale: Complex gestures and handheld devices such as controllers may be impractical during motion.

6. **Is the expected user population likely to have accessibility needs (e.g., low vision, motor impairments)?**
   Rationale: Different impairments affect the usability of input types. Accessibility-conscious design improves inclusivity.

7. **Is precision or speed of input critical (e.g., emergency response, surgical assistance)?**
   Rationale: Some domains demand quick, error-free inputs that are less tolerant of ambiguity.

8. **Does the domain have physical constraints (e.g., gloves, protective gear, low or extremely bright light)?**
   Rationale: Certain input types (touch, visual gesture) may be hindered by the environment or equipment.

9. **Is cost, battery life, or hardware complexity a major concern?**
   Rationale: Simpler input modalities can reduce device cost and power consumption, which is crucial in field or scalable applications.

Table 3.3.: Input modalities and interaction types

Input type	**Input Modalities**	**Interaction type**	**Advantages**	**Challenges/Drawbacks**
External devices	Controller	Gesture, raycast, click	Intuitive, precise, haptic	Sterilization, hand occupation, extra cost and bulky hardware
	Clicker/ button	click, hold, release	Intuitive, precise, haptic, subtle and non-intrusive, well suited for socially sensitive environments	Sterilization, hand occupation, limited functionality
	Foot pedals	Tap, hold, release	Moderately intuitive, hands-free, good in gloved or sterile conditions	Limited functionality
No external devices	Full body	Gesture	Optimal for activity inspired system design	Requires externals or on body sensing
	Head (captured via IMU/head-mounted camera)	Gaze, Gesture	Intuitive	Limited functionality, not optimal for repetitive use
	Eye (Eye tracking)	Gaze, Dwell	Intuitive	Limited functionality (should be combined with another modality), requires user attention and causes visual distraction
	Voice (captured via microphone)	Speech commands	Precise, hands-free, useful for users with motor impairments	Memorization issues, not optimal for object manipulation, not optimal for noisy or privacy sensitive environments

Input Type	Input Modalities	Interaction type	Advantages	Challenges/Drawbacks
No external devices		Natural speech	Intuitive, hands-free, useful for users with motor impairments	Not optimal for object manipulation, need context aware system design or combined with another modality (e.g. Eye gaze), not optimal for noisy or privacy sensitive environments
	Hand (captured via head-mounted camera)	Mid-air touch, raycast	Precise, intuitive	Hand occupation, challenging for far field interaction if not combined with other modalities (e.g. eye/head gaze, voice), difficult to use when walking or moving, prone to error when using gloves or equipment
		Mid-air Gesture	Extended functionality range, optimal for object manipulation, no sterilization issue	Hand occupation, learning curve, not ideal for confined spaces, difficult to use when walking or moving, prone to error when using gloves or equipment
	Foot (captured via external or head-mounted camera, wearable sensing)	Mid-air touch, mid-air gesture	Hands-free, useful for seated tasks, viable in constrained setup	Difficult to perform, Limited functionality, not optimal when walking, difficult to detect due to occlusion and low visibility
	Brain (captured via EEG)		Interaction without physical movement, suitable for users with severe disabilities	Prone to noise, hard to interpret, complex setup, needs further research

While user setup and environmental constraints are critical in choosing input modalities for wearable AR, the intended functionality of the interaction is equally influential. Different tasks place different demands on precision, speed, intuitiveness, and feedback, which may render some modalities more effective than others. For example, text entry often requires high precision and rapid input, making voice dictation or external keyboard integration more suitable than mid-air gestures. In contrast, object selection in 2D space can be efficiently handled using touch or gaze-dwell interactions.

More complex operations, such as manipulating a 3D object in space (e.g., rotating or resizing a model in surgical planning), may demand a combination of modalities. Here, relying solely on voice, e.g., saying "rotate left 20 degrees", can be slow and cumbersome. However, combining voice (for command input) and gaze (for selecting or anchoring the object) enables both precision and efficiency. Similarly, a combination of hand gestures (for motion) and haptic feedback (for collision or limit indication) can make 3D manipulation far more intuitive and controllable.

This highlights the importance of multimodal interaction design tailored to specific functions, not just user profiles. Designing AR systems with layered modalities allows the user to switch seamlessly or use modalities in tandem, resulting in a more natural and effective interaction experience.

Table 3.4.: List of functionality type in in augmented reality head-mounted displays and corresponding suitable interaction types.

Functionality Type	**Suitable Interaction Type**	**Example Functionality**
Selection (click)	Touch, Gesture, Gaze / Raycast, speech commands, natural speech	Highlight, identifying focused object
Confirmation (double-click)	Touch, Gesture, Gaze / Raycast, speech commands, natural speech	System control (UI/Menu)
Navigation	Gesture, speech commands, natural speech	Pan, scroll, zoom in/out
Object Manipulation	Gesture, speech commands (very limited), natural speech (challenging)	Scale, rotate, deform, move
Object Management	Gesture, speech commands, natural speech	Grouping, stacking, delete, duplicate

3.2.3. Designing Outputs: Managing Cognitive Load Through Feedback

The choice of output modalities or feedback plays a critical role in determining user cognitive load and the overall usability of a system. These modalities directly influence how effectively users can perceive, interpret, and act upon the information provided by the system. AR systems utilize a variety of output modalities to engage users across multiple sensory channels, thereby enhancing immersion, interaction, and usability. These modalities can be categorized as follows:

1. **Visual Output**: This is the most common modality in AR and includes:
   - Head-mounted displays (HMDs) and smart glasses
   - 2D and 3D overlays on the physical environment
   - Spatial mapping and object anchoring
   - Visual highlights, color cues, and animations
2. **Auditory Output**: Sound is used to convey information, direct attention, and enrich context through:
   - Spatialized audio (3D sound positioning)
   - Verbal instructions (text-to-speech)
   - Audio alerts and notifications
   - Ambient or environmental sound effects
3. **Haptic Output**: Tactile feedback enhances physical interaction and can include:
   - Wearable devices (e.g., gloves, suits, wristbands)
   - Force feedback (e.g., joysticks, styluses)
   - Vibration cues from mobile devices or AR controllers
4. **Olfactory Output** (less common): Scent emitters simulate environmental smells for immersive training or therapeutic applications.
5. **Thermal Output** (experimental): Devices that simulate temperature changes to enhance sensory realism and feedback.
6. **Tactile/Air-based Output**: Systems that generate touch-like sensations using air jets or ultrasonic waves, useful for contact-free or sterile environments.

While the selection of output modalities is often constrained by the technical capabilities of the device in use, it is equally important to consider the specific needs of the user as well as the environmental context in which the system operates.

In educational environments, for instance, the integration of multiple sensory outputs (e.g., visual and auditory cues) can facilitate deeper engagement

and improve knowledge retention through multisensory learning. Conversely, in high-stakes or precision-critical scenarios such as surgery, excessive sensory input may overwhelm or distract the user, potentially compromising safety and performance. In such cases, minimal and targeted output modalities, such as discrete auditory alerts or focused visual overlays, are often more effective.

Moreover, certain output modalities may be impractical or even impossible to implement in specific environments. For example, while haptic feedback can enhance spatial awareness and tactile interaction in many applications, it may be unsuitable in sterile or surgical settings due to hygiene and sterilization requirements. These constraints necessitate careful evaluation of both the benefits and the feasibility of each modality in context-specific use cases.

Ultimately, the design of output modalities should be guided by a balanced consideration of technological feasibility, user cognitive capacity, and environmental limitations to ensure optimal usability and user experience. Table 3.5 lists a series of questions along with their associated implications, intended to aid in identifying appropriate output modalities based on user needs and environmental constraints.

Table 3.5.: Framework for Selecting Output Modalities in AR Systems

Modality Area	Question	Implications
General	What is the user's goal in this AR experience?	Aligns modality with task needs (e.g., visual for inspection, audio for navigation).
	What information is critical vs. supplementary?	Use robust modalities for critical info; less intrusive ones (e.g., haptics) for supplementary data.
	Is real-time interaction or response required?	Low-latency modalities (e.g., visual, haptic) are preferable for real-time feedback.
	What is the cognitive load of the user at this point in the task?	Avoid overloading; distribute information across underutilized senses.
Visual	Will the user's visual attention already be heavily engaged?	Avoid additional visual input; consider audio or haptic alternatives.
	Can visual cues be clearly distinguished in the current environment?	If not, reinforce or replace with other modalities.
	Is ambient lighting sufficient for overlays?	Poor lighting requires modality adaptation (e.g., high-contrast visuals or audio alerts).
	Will visual clutter overwhelm or distract the user?	Use minimal, context-aware overlays to prevent information overload.
Auditory	Is the environment noisy or quiet?	In noisy areas, audio may be ineffective; use visual/haptic reinforcement.
	Will auditory cues interfere with communication?	Avoid verbal outputs if users are speaking/listening; consider tones or visual aids.

Modality Area	Question	Implications
	Can spatial audio direct attention?	3D audio can improve spatial awareness without visual load.
	Are headphones or some sort of near-ear speaker feasible or appropriate?	If not, audio may need to be discreet or replaced with other modalities.
Haptic	Is the user able to wear haptic devices comfortably?	Discomfort reduces usability; favor lightweight or alternative feedback.
	Are there hygiene or sterility requirements?	Haptic devices may be restricted in sterile environments; consider non-contact feedback.
	Will haptics enhance motor skills or awareness?	Use for reinforcing touch-based tasks or spatial guidance.
	Is latency in haptic feedback acceptable?	High-latency haptics can disrupt precision tasks; ensure responsiveness.
Cognitive Load & UX	Does the modality reduce or increase mental effort?	Choose modalities that simplify understanding and reduce workload.
	Can it offload cognitive work from overloaded senses?	Distribute load through multimodal balance (e.g., audio vs. visual).
	How experienced is the user with this modality?	Novices may prefer familiar outputs; experts may handle more complex inputs.
Environment & Technical	What are the physical conditions of the environment?	Select resilient modalities (e.g., visual in quiet areas, audio in low-light).
	What are the hardware limitations?	Optimize for device constraints (e.g., battery, resolution, audio quality).
	Is mobility required?	Avoid bulky or tethered modalities in mobile scenarios.

Chapter 4

Procedural and Motor Skill Acquisition

Contents

Medical training systems often rely heavily on the continuous presence and guidance of a qualified trainer to deliver educational content. This includes instruction on procedural guidelines, proper use of medical equipment, and crucially, the ability to provide timely feedback and interventions during practice. While this instructor-led model supports knowledge acquisition and skill development, it presents notable limitations in terms of cost, time efficiency,

and scalability. These constraints often result in reduced one-on-one instructional time for each student.

In the medical domain, where the quality of healthcare education directly impacts patient outcomes, such limitations represent a significant barrier. The diversity in learners' skill levels, learning styles, and confidence further compounds this issue. A standardized, time-limited training approach may fail to accommodate individual learning needs, potentially leading to insufficient procedural competence, reduced self-efficacy among healthcare professionals, and ultimately, compromised patient care.

This chapter investigates the role of augmented reality (AR) in supporting procedural and motor skill acquisition within the healthcare domain. Using cardiopulmonary resuscitation (CPR) training as a case study, an essential component of emergency care, it examines how AR-based systems can enhance learning outcomes for both healthcare professionals and lay responders.

The content of this Chapter covers the following publication: .

> Javaheri, Hamraz, et al. "RescuAR: A Self-Directed Augmented Reality System for Cardiopulmonary Resuscitation Training." International Conference on Pervasive Computing Technologies for Healthcare. Cham: Springer Nature Switzerland, 2023.

4.1. Background

Cardiopulmonary resuscitation (CPR) plays a crucial role in saving lives during cardiac arrest, a medical emergency with high mortality rates. The timely and effective administration of CPR significantly increases the chances of survival [97]. Therefore, it is vital to ensure that individuals are well-trained in this life-saving technique. Traditional CPR training methods, such as classroom-based instruction and mannequin practice, have been the cornerstone of CPR education for decades. However, these methods have shown limitations in terms of skill acquisition, retention, and providing immediate feedback on performance.

One prominent challenge in traditional CPR training lies in the gap between knowledge acquisition and practical application. While learners may grasp the theoretical concepts, transferring that knowledge into effective hands-on performance can be challenging [31]. Research has shown that individuals often struggle to translate their theoretical knowledge into practical skills when faced with high-stress situations, such as cardiac arrest scenarios [150].

Furthermore, the ability to receive real-time feedback on performance during CPR training is critical for learners to correct errors, refine their technique, and build confidence. Immediate feedback allows learners to adjust their actions, ensuring the application of correct chest compression (CC) depth, rate, and recoil. However, traditional methods of CPR training often lack the means to provide instantaneous and accurate feedback, leaving learners uncertain about their proficiency and limiting their ability to improve their skills effectively.

To address these challenges and bridge the gap in CPR education, the combination of AR technology with sensing modalities emerges as a promising solution. AR integrates virtual elements into the real-world environment, offering learners an immersive and interactive training experience, while sensing modalities provide an opportunity for the integration of real-time feedback. By leveraging computer vision and sensing technologies, these systems can analyze the learner's movements and provide immediate feedback on the accuracy and effectiveness of their CPR technique. This real-time feedback enables the learner to make adjustments on the spot, improving the quality and consistency of their compressions. This approach enhances skill acquisition, retention, and performance by overlaying instructional guidance and visual cues onto the learner's view of a CPR scenario [8, 98, 278].

While existing studies on AR-based CPR training have shown promise in enhancing training experiences and outcomes [17, 110, 114, 127, 129, 149], there is still a need for further research to fully explore the effectiveness of AR in addressing the current limitations of traditional CPR training methods. Although AR has demonstrated potential in providing immersive and interactive training environments [77, 119, 129, 143, 222, 279], its specific impact on skill acquisition, retention, and performance during CPR training is an area that requires more investigation. Understanding the potential benefits and challenges associated with incorporating AR technology into CPR training can inform the development of evidence-based guidelines and best practices for its further integration.

4.2. Methodology

Upon reviewing the existing literature, we identified a gap in the research regarding the detailed exploration of the application design process for AR-based CPR training tools, as well as the limited availability of quantitative data for a comprehensive comparison between these tools and traditional CPR training methods. Motivated by these gaps in the existing literature, our study aimed to fill this research void by answering the following research questions:

RQ1: What are the required design features and functionalities for an AR-based self-directed CPR teaching and training tool?
We answer this research question by conducting an extensive design survey among healthcare professionals and investigating their professional opinion on the required features and specifications for such a training tool.
RQ2: Are the performance results of the designed system comparable with traditional teaching methods?
Our objective was to address this question by conducting an experimental trial where we compared the CPR performances of participants who used the designed system to learn and practice the CPR routine with participants who underwent traditional teaching. We measured

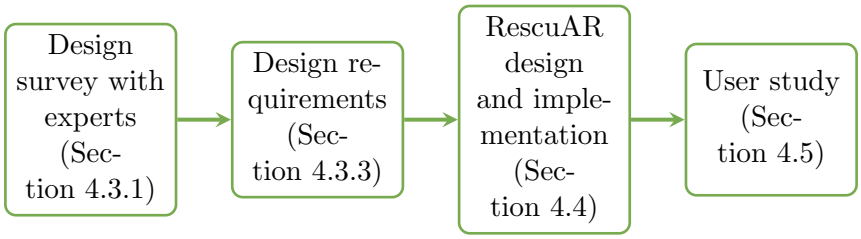


Figure 4.1.: The study flow process starting with the specification of design requirements based on expert survey findings and later evaluation of the designed system in a user study

their performances before and after teaching sessions in terms of correct depth, correct frequency, and overall effective CPR (correct depth and correct frequency) to develop a better understanding of the efficiency of the designed system.

We employed a mixed-method design consisting of three consecutive steps: data collection for design requirements, system design and implementation, and system evaluation (Figure 4.1). Initially, a survey was conducted among a group of health professionals to gather valuable insights and identify the design requirements and constraints for an AR-based self-directed CPR teaching and training tool. The survey aimed to understand the preferences, needs, and expectations of professionals regarding CPR training, as well as their perceptions of the potential benefits and challenges of using such an interactive system.

Based on the insights and design requirements identified through the survey, RescuAR, an AR CPR self-training tool, was conceptualized and designed. The primary goal of RescuAR was to enable users to learn and improve their CPR performance skills by providing them with interactive features and real-time data visualization.

To assess the effectiveness of RescuAR, a user study was conducted involving nurse students and laypeople. The participants were randomly assigned to either the experimental group, which utilized RescuAR for CPR training, or the control group, which underwent traditional teaching methods. The CPR performances of participants were evaluated before and after the training sessions, specifically focusing on parameters such as compression depth and frequency. As various studies demonstrated the importance of hands-only CPR in bystanders [20, 154], this study's primary objective was to evaluate participants' CPR performances specifically regarding hands-only CC, and any effect related to breath technique was neglected.

4.3. Design Requirements for Self-directed AR-based CPR Teaching and Training Tool

To inform the design of our interactive CPR teaching and training tool, we conducted a survey among experienced healthcare professionals. All professionals were familiarized with the AR device capabilities that will be used for the study prior to conducting the survey. The purpose of the survey was to identify the key requirements (RQ1) for developing an AR-based self-directed learning system that could effectively support CPR education. By gathering insights from these professionals, we aimed to create a tool that addresses the specific needs and challenges of learners in the context of CPR training.

4.3.1. Design Survey

The survey employed a combination of open-ended and multiple-choice questions. To provide a more structured approach, the questions were categorized into five categories: User needs and requirements, user interface and design, instructional content, interaction and feedback, and miscellaneous.

While the questions under the user needs and requirements category focused on the participants' opinions on the features and functionalities that are considered essential for an AR-based CPR training app, the user interface and design questions focused more on the appearance and aesthetic aspects of the app.

In the instructional content section, several questions were asked to gather insights regarding the content, flow, and delivery mode of the teaching materials. Participants were asked to provide feedback on the clarity and comprehensiveness of the instructional content, as well as their preferences for the organization and sequence of the material. Additionally, participants were asked about their preferred delivery modes, such as text, images, videos, or interactive elements. Furthermore, more detailed questions focused on finding the most effective teaching methods for acquiring correct CPR skills, such as correct CC depth and frequency, along with suggestions for optimization.

Under the interaction and feedback section, participants were asked to express their preferences for input methods, such as voice commands, button interactions, and gestures. Additionally, participants were asked to indicate their preferred methods of real-time feedback from the app, whether visual, audio, or haptic.

Lastly, the miscellaneous section encompassed more general questions, allowing participants to share their opinions on the gamification aspects of the AR-based CPR training app. Additionally, participants were given the opportunity to provide any overall suggestions or feedback they deemed relevant to the development of the app.

By capturing diverse perspectives from healthcare professionals, the survey played a crucial role in informing the design and development of an effective AR-based CPR training app.

Table 4.1.: The participants' characteristic distributions. IQR: interquartile range

Characteristics	
Age (years), Median (IQR)	39, (32-48)
Female	2
Male	9
Non-Binary	0
Experience in healthcare (years), Median (IRQ)	20 (14-23)
Tried or Familiar with AR (yes)	6 (54%)

4.3.2. Participants

A survey was conducted among 11 healthcare professionals. The participants' demographic characteristics, including their professional backgrounds and relevant experience, are summarized in Table 4.1.

4.3.3. Survey Findings

The open-ended questions in the survey were analyzed using qualitative data analysis methods to evaluate the responses and gain deeper insights into participants' perspectives. The analysis process involved systematically reviewing the open-ended responses, coding them for key themes and patterns, and organizing the data into meaningful categories. Through this approach, we were able to identify common themes, explore variations in participants' experiences and opinions, and gain a comprehensive understanding of the topics under investigation. The qualitative analysis provided valuable qualitative data that complemented the quantitative findings, allowing for a richer and more nuanced interpretation of the survey results. These results highlighted the essential design elements and interaction methods that were considered crucial for an effective CPR training application based on the professionals' expertise and daily experiences in the healthcare field.

Instructional Content, Routine, and Design Elements

The survey respondents provided valuable insights and reached a consensus on several key aspects of the CPR training application design. They emphasized the importance of incorporating a combination of audio, scripted text, and a virtual human teacher within the app to effectively deliver theoretical materials and provide personalized guidance and support. Additionally, participants strongly favored the use of animated human avatars as a demonstration tool for CC and breath techniques compared to other methods, such as 2D videos of real persons performing CPR or audio instructions and verbal cues. The dynamic and interactive nature of animated avatars was perceived as more engaging and effective in conveying the correct techniques for CPR training, as it would enable students to observe the correct action in 3D. In terms of the teaching routine, participants suggested that it should begin with a theoretical

representation of the concepts, followed by hands-on practice. They further recommended that the teaching of compression depth and frequency should be initially conducted separately before combining them in the training sessions. Lastly, participants highlighted the significance of enabling repetitive theoretical and practical sessions. They emphasized the importance of incorporating functionality to navigate through the different stages of the training. This feature would provide users with the ability to easily access previous and next steps, facilitating repetitive practice sessions. The opportunity for repetitive practice allows learners to review and reinforce their CPR skills, ultimately improving their performance and confidence.

Input and Interaction

Regarding input and interaction, the survey respondents expressed a strong preference for the integration of interaction methods through physical buttons or verbal communication via voice commands. Both methods were considered intuitive and practical for engaging with the app during training sessions. The respondents emphasized the advantages of voice commands, as they enable users to perform actions without the need for manual input, allowing them to focus on performing CPR on the manikin. This hands-free interaction was seen as a convenient and efficient way to engage with the app. On the other hand, physical buttons were identified as a suitable option for environments with high noise levels or crowded settings. Furthermore, participants expressed the belief that using mid-air hand gestures or touch (using virtual buttons) for interaction would not be suitable for this task, as they would require users to have prior knowledge of how to interact with such input methods.

Methods for Frequency and Depth Acquisition

The survey respondents highlighted the crucial role of immediate and real-time feedback during hands-on training sessions for acquiring the correct frequency and depth in CPR. Among various types of feedback, the combination of audio and visual feedback emerged as the most preferred method for both frequency and depth acquisition.

In terms of frequency acquisition, healthcare professionals recommended the inclusion of audio elements to facilitate proper timing during CPR. They expressed the belief that incorporating the iconic “Stayin’ Alive” sound, which aligns with the required rhythm for CPR, would enhance the training experience. However, to promote stronger muscle memory and improve precision, professionals also suggested the inclusion of a metronome sound.

For correct depth acquisition, the respondents emphasized the importance of incorporating a simple graphical visualization, similar to the devices used in traditional teaching setups, to provide visual cues on the depth of CC. Participants pointed out the importance of optimizing the orientation and location of the visualization within the user’s field of view to ensure clarity and accuracy during CPR training. They recommended that the depth display be positioned near the manikin’s chest within the user’s field of view, in a fixed

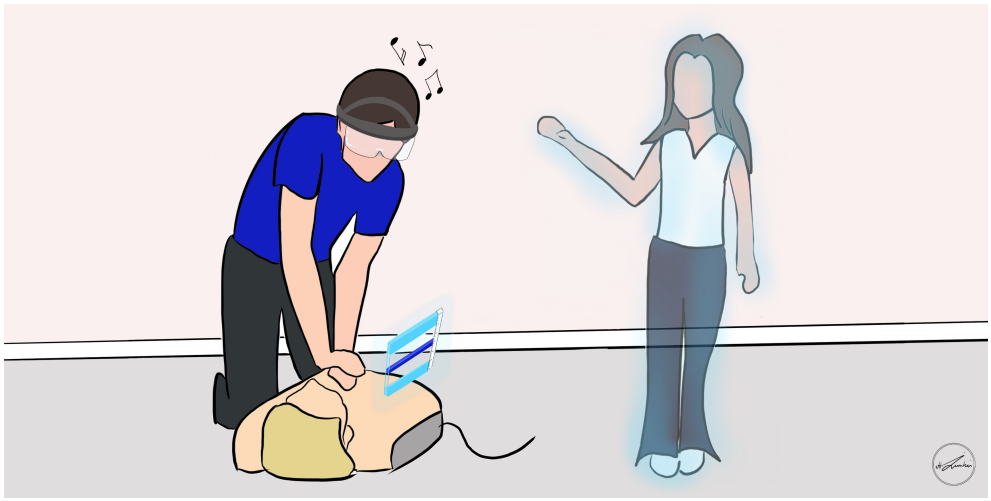


Figure 4.2.: A sketch demonstrating the conceptual design of RescuAR.

position to maintain proper posture during performance. However, considering that CPR is a highly physical activity and individuals may have different preferences, participants suggested that it would be beneficial if the position of the depth display could also be adjustable by the user based on their personal preference. This customization feature would allow users to optimize their viewing experience and ensure optimal training engagement.

4.4. RescuAR System Design and Implementation

Based on survey findings, a concept design was sketched (Figure 4.2), and following that an AR-based system prototype was developed with the aim of creating an immersive multi-sensory CPR teaching and training tool, incorporating audio, visual, and tactile elements, to enhance the learning experience and real-time feedback to foster a comprehensive understanding of the CPR technique.

The application was designed and developed using the Unity 3D game engine [258]. Microsoft HoloLens [179] was used as a wearable AR device to run the application.

The system consisted of two main parts: the RescuAR application and a standard CPR manikin [146] covered with custom pressure sensors. The RescuAR application utilized real-time data streams from the CPR manikin to provide feedback on compression rate, depth, and recoil. A local wireless communication scheme facilitated seamless interaction between the manikin and the application, enabling accurate and immediate feedback on CPR performance.

4.4.1. RescuAR Application

The formulation of the CPR training app's prototype design was based on the results derived from both quantitative and qualitative data collected through

the survey. According to the survey participants, a two-phase delivery of teaching material was found to be appropriate. They recognized that understanding the underlying principles and concepts of CPR is crucial in order to perform the techniques accurately and confidently. Hence, the application routine was divided into two distinct phases: Theory and practice. During the initial theory phase, the application focused on imparting the fundamental principles of CPR. The subsequent practice phase emphasized hands-on practical training. The flow chart of the application routine is presented in Figure 4.3.

Given that the primary objective of this study was to evaluate participants' performance specifically regarding CC, the instructions pertaining to rescue breaths were exclusively included in the theoretical section and not incorporated into the practical training. This deliberate decision allowed for a more targeted assessment of participants' proficiency in CC, aligning with the study's primary focus.

Theory Phase

During the theoretical phase, users were introduced to the fundamentals of CPR (airway, breathing, and circulation techniques) based on European resuscitation council guidelines [196] through an engaging virtual teaching experience. A virtual teacher avatar was designed following the survey results and served as the guide, delivering the essential CPR basics both audibly and visually. To enhance comprehension, the information was presented not only through the avatar's spoken instructions but also as easily understandable written scripts. This method was deemed to be the most suitable approach for this application by survey participants. The communication between the virtual teacher and the participant was facilitated using voice commands. The virtual teacher provided the voice commands at the end of the conversation to reduce the need for memorization. For example, the virtual teacher would say, "If you feel like you need further practice, you can let me know at any time. Just use the word 'repeat!"

To ensure a clear understanding of proper technique, animated 3D avatars were employed to visually demonstrate the correct posture and hand positioning during CPR (Figure 4.4). By utilizing 3D virtual models and realistic placement within the room, participants were able to walk around the models and observe the CPR technique from different angles, providing a more immersive and interactive 3D training experience that cannot be achieved using 2D displays. By combining audio instruction, written scripts, and animated avatars, the theoretical phase of the training aimed to maximize participant comprehension and knowledge acquisition.

The application commenced with a calibration phase, which involved scanning the environment and room using the capabilities of the HoloLens and Mixed Reality Toolkit (MRTK) to obtain a spatial mesh of the environment. We utilized plane-finding methods to detect suitable surfaces for the placement of virtual teacher avatars and other demonstrative avatars. Colliders were added to the room mesh to create a more realistic movement area for the

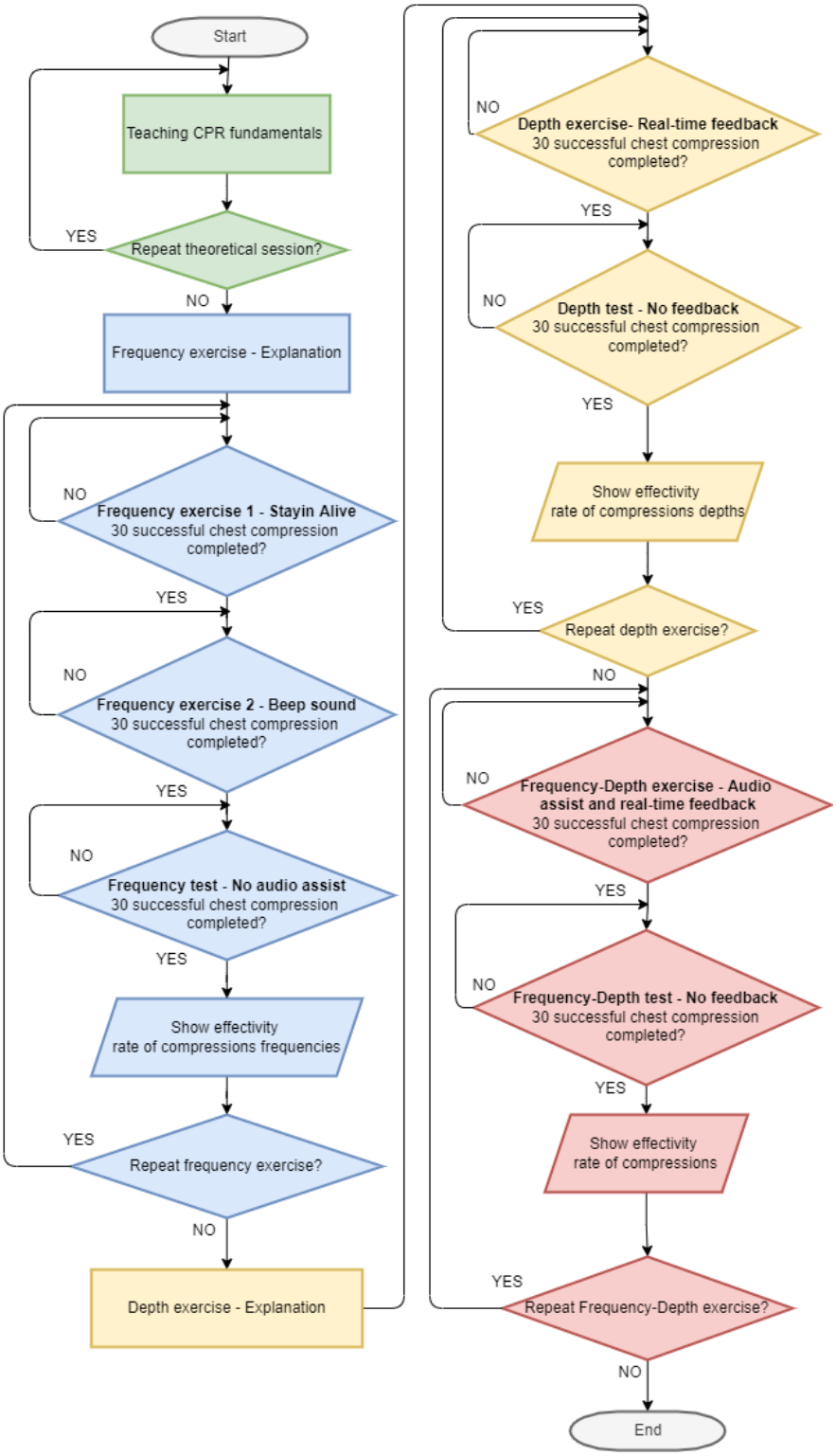
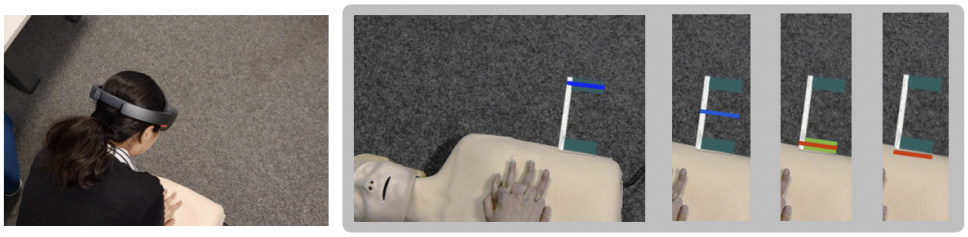


Figure 4.3.: The flowchart explaining the routine of the RescuAR CPR tutorial and training application

(a)



(b)

Figure 4.4.: Third-person and first-person (in-app) views of the RescuAR application. (a) Theoretical phase. (b) Practical phase

teacher avatar.

Practice Phase

The second phase of the application focused on practical training to cultivate the essential muscle memory required for CPR proficiency. Based on the feedback from survey participants, separate hands-on practice addressing each essential criterion of CPR was emphasized as important. As a result, two key criteria, frequency and depth, were targeted individually and then combined to ensure comprehensive learning.

To teach the correct frequency of CC, two distinct audio cues were employed. Initially, the iconic "Stayin' Alive" song, known for its rhythm matching the recommended CPR procedure (104 bpm), was utilized to encourage participants to perform CC in sync with the song's beats. Subsequently, a metronome sound with the same frequency (104 bpm) was introduced to enhance learner focus and synchronization.

Regarding CC depth, a trial-and-error approach was adopted to instill the appropriate technique. To align with the preferences of professionals who recognized the benefits of using a design similar to standard commercial devices, real-time feedback on compression depth was provided through a visual depth panel resembling CPRmeter devices [175]. This allowed learners to assess their performance and make adjustments accordingly (Figure 4.4). The application allowed for the repositioning of the virtual depth feedback panel using the manipulation gesture of HoloLens. To address the shaking effect that occurred during CPR performance, we implemented a functionality to fixate the depth

feedback panel to the corner of the field of view. Participants could enable or disable this feature using a voice command. To eliminate the need and urge to look at the teacher avatar during chest compressions, which could be affected by the shaking effect, the visual rendering of the avatar was disabled during the practice session and reappeared after the completion of the stage. This effect was also explained by the teacher to the user to avoid any confusion. Additionally, a click sound reminiscent of the standard CPR dummy's internal clicker (which had been previously removed to avoid confusion) indicated the moment when the correct depth was achieved.

Upon completion of training for both frequency and depth criteria, a final session provided an opportunity for learners to combine both elements, aiming to perform CC with the correct frequency and depth in a synchronized manner. This comprehensive training approach equipped participants with the necessary skills to deliver effective CC during CPR.

4.4.2. CPR Manikin:

To acquire information such as depth, frequency, and pressurized position of the CC on the manikin, a single FSR (Force-Sensing Resistor) was used. The sensor was attached beneath the skin layer of the CPR manikin's chest plate, precisely positioned in the center of the chest where the hands should be placed during correct CPR (Figure 4.5). To facilitate seamless data acquisition and control, the pressure sensors were directly connected and managed by an Arduino Pro Mini board [12]. The analog signals generated by the embedded sensors were transformed into digital signals through the Arduino platform. Subsequently, these digital signals originating from the sensors were transmitted via a serial port to a local computer for further utilization and analysis.

To calculate the depth of chest compressions, we adopted the method proposed by Tsou *et al.* in their study [256]. This method utilizes Hooke's law to calculate the depth based on the pressure applied to the chest spring. We assumed the spring inside the dummy to be a linear spring due to the same diameter along its entire length. We converted the FSR readings to Newton to calculate the spring constant. Before the experiment, we measured the spring constant using the FSR force value and the spring displacement using the following formula:

$$x = \frac{F}{k}$$

Where F is the force applied to the compression spring (reading from the pressure sensor converted to Newtons, N), k is the spring constant ($\frac{N}{m}$), and x is the displacement of the spring (m).

The accuracy of the employed method for calculating the depth of CC was verified against a commercial CPRmeter device [175] to ensure the validity of the measurements.

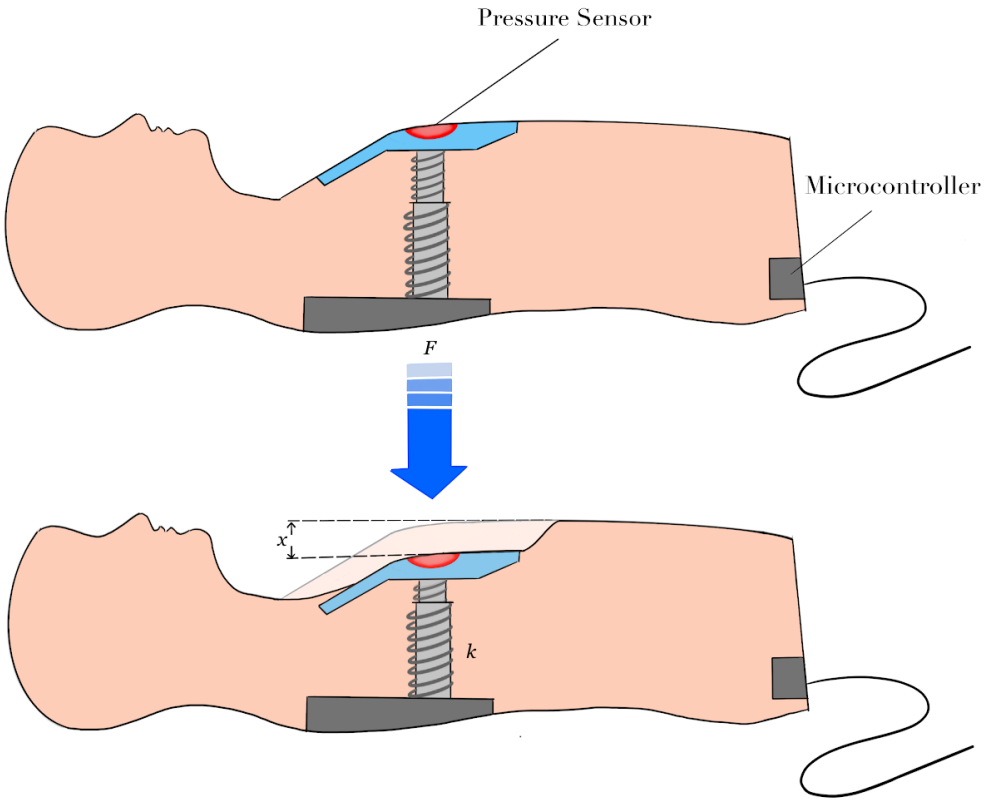


Figure 4.5.: Placement of Pressure sensors and controller inside the manikin. F = Applied force to the spring (N), k = spring constant ($\frac{N}{m}$), x = displacement of the spring (m)

4.5. User Study and System Evaluation

To evaluate the effectiveness of our designed RescuAR system, we conducted a user experiment. The study took place in two different experiment centers, Southampton University, Southampton, UK, and German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany. The primary objective of this experiment was to assess and compare the CPR performance of participants who used the RescuAR system with those who underwent traditional CPR training methods (RQ2).

4.5.1. Study Design

The study utilized a randomized controlled trial design to investigate the effectiveness of different teaching methods for CPR training. Participants were randomly assigned to either the experimental or the control group.

Experimental Group

The experimental group received a self-directed CPR teaching and training session using the RescuAR system (Application and the CPR manikin). The teaching session began by calibrating the HoloLens for each individual and starting the application. Each participant completed the teaching and training session without any time limit.

Control Group

The control group underwent a traditional CPR teaching and training session, which involved classroom-based instruction and practice using the CPR manikin. The same manikin as the experimental group was used in the control group to avoid any biases. A certified teacher was recruited to provide essential information on performing CPR based on European Resuscitation Council guidelines [196]. The session began with a theoretical introduction to the airway, breathing, and circulation techniques, followed by a demonstration from the teacher on the correct CPR procedure using the CPR manikin. Participants were then given the opportunity to practice CPR on the same manikin under the observation of the teacher. The teacher interrupted and gave feedback whenever they felt essential. Throughout the session, participants were encouraged to ask questions, repeat the training, and perform additional CC cycles if they desired. No time constraints were applied during the teaching and training sessions. To minimize bias between the groups, the training sessions were conducted on an individual basis.

Both groups received instruction on airway, breathing, and circulation techniques. However, during the data recording session, participants were only asked to perform chest compressions, as the main focus of our study was to evaluate the quality of chest compressions in terms of depth and frequency while neglecting the potential effects of the breath technique. No device or extra feedback method was used during data recording. The performances of participants were assessed before and after the study to measure their baseline CPR skills and the improvements achieved through the assigned teaching method. By comparing the performance improvements between the experimental and control groups, the study aimed to evaluate the impact of the RescuAR system on enhancing CPR skills.

4.5.2. Study Protocol

This protocol was reviewed and approved by the ethical committee of Southampton University. All of the participants were informed about being free to participate in the research and the nondisclosure of personal information. They all agreed and signed written informed consent. Upon written consent, the experiment protocol was performed in five ordered stages: (1) Demographic survey, (2) baseline CPR recording, (3) randomized group assignment, (4) teaching and training session, and (5) post-training CPR recording. The survey collected demographic and characteristic information, including age,

gender, experience background in CPR, and familiarity with wearable AR devices. After completion of the survey, participants were asked to perform two cycles of 1-minute hands-only CPR. Participants rested for at least two minutes between each cycle. Later, all participants were randomly divided into two experimental groups to receive teaching and training sessions. After the teaching session, participants rested for at least 10 minutes to avoid the effect of tiredness on their performance. Later, they performed another two cycles of 1-minute hands-only CPR (without the use of any feedback device or extra help) with at least two minutes rest between each cycle. As various studies demonstrated the importance of hands-only CPR in bystanders [20, 154], this study only focused on the evaluation of hands-only CPR performance, and any effect of performing rescue breath was neglected.

4.5.3. Participants

A total of 44 persons volunteered to participate in this study. Among those, 43 persons' data were included, and one person's data were excluded from the study due to data corruption. After exclusion, the experimental group consisted of 22 volunteer participants, and the control group contained 21 participants. The participants were nurse students and laypeople who were randomly assigned to study groups. The nurse students were recruited by Southampton University, and laypeople were recruited by DFKI. All participants were required to be aged above 18. Characteristics of participants are presented in Table 4.2.

Table 4.2.: The participants' characteristic distributions. IQR: interquartile range; With prior knowledge, who is CPR certified and/or completed a first aid course

Characteristics	**Control**	**Experimental**
Age (years), Median (IQR)	22, (20-24)	21, (19-24)
Female	5 (23%)	9 (41%)
Male	16 (77%)	13 (59%)
Non-binary	0 (0%)	0 (0%)
Nurse Student	9 (43%)	9 (41%)
Laypeople	12 (57%)	13 (59%)
With prior knowledge	8 (38%)	11 (50%)
Tried or Familiar with AR (yes)	2 (9%)	1(4%)

4.5.4. Data Collection and Analysis

Over two 1-minute cycles, two CC measurements were recorded before and after the teaching session using the same sensor-equipped CPR manikin described in Section 4.4.2 with a sampling rate of 100 Hz. The effective CPR performances of the participants were analyzed according to the latest evidence-

based guidelines for resuscitation officially published by the European Resuscitation Council [208]. These guidelines suggest an effective CPR as follows:

- CC in a frequency of at least 100/min but not exceeding 120/min.
- CC with a depth of at least 5 cm but not exceeding 6 cm

In this study, the percentage of effective CPR performance of a participant was calculated based on the percentage of the time that the participant complied with all the above-mentioned guidelines at the same time while doing CPR.

Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 22.0 [50]. Numerical data were presented as means ± standard deviations. To compare the improvement in the two groups before and after teaching sessions, a paired t-test was used. To compare the outcomes between two study groups, the chi-square test (for categorical data) and Student's t-test (for numerical data) were used. A two-sided p-value less than 0.05 was considered significant in all analyses.

4.5.5. Results

RescuAR Evaluation

The participants in the experimental group showed significant improvements in performing correct depth and frequency. While the calculated p-value for CC depth improvement was 0.003, the p-value for frequency improvement was below 0.0001. A comparison between both criteria before the experiment session showed that more participants had issues finding the correct frequency (39.8% ± 36.9% correct frequency) than the correct depth (58.7% ± 39.8% correct depth). Even though for most of the participants, both depth and frequency were improved after the tutorial and training session, the frequency improvement rate was higher (73.8% ± 28.2%) than the improvement rate in CC depth (84.3% ± 23.2%).

Moreover, the results showed that the overall effective CPR performance of the experimental group significantly increased ($p < 0.0001$) after the teaching session with RescuAR. While the mean of participants' performance was 23.3% ± 27.2% before the teaching session, it was improved to 61.3% ± 27.4% after training with the proposed system.

Experimental Group vs. Control Group

The study analysis showed no significant difference regarding characteristics distribution between the two groups concerning sex ($p = 0.332$) and knowledge backgrounds ($p = 0.543$) of the participants. Based on the findings, RescuAR helped to achieve higher effective CPR rates compared to traditional teaching. Analyzing the final performances of both groups' participants, it was observed that, even though both groups had no significant difference in doing effective CPR before the teaching session (control = 20.3% ± 25.4%, experimental = 23.3% ± 27.2%, $p = 0.594$), the experimental group performed significantly

better than the control group after the teaching session (control = 40.4% ± 28.5, experimental = 61.3% ± 27.4%, $p = 0.019$) (Figure 4.6).

Moreover, performance degradation only occurred in some of the control group's participants, and all participants in the experimental group showed improvement after the teaching session (Figure 4.7).

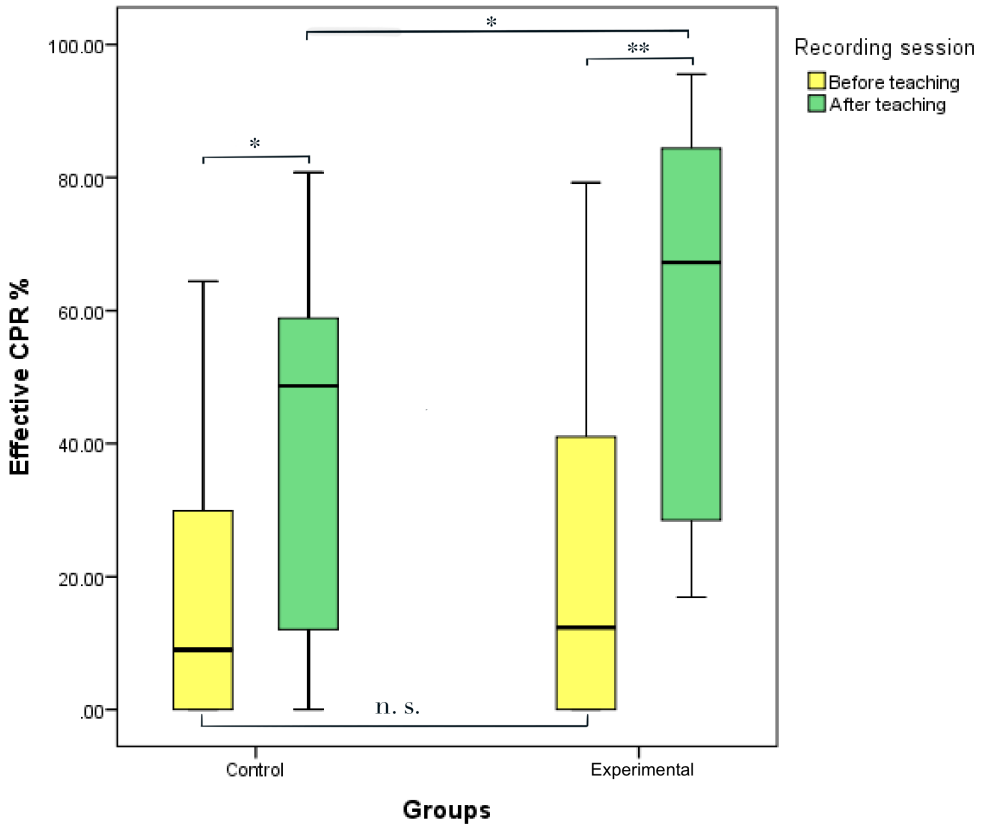


Figure 4.6.: The figure demonstrates the statistical analysis of control and experimental groups - n.s. = no significant difference between indicated groups ($p < 0.05$) * = Significant difference between indicated groups ($p < 0.05$), ** = Significant difference between indicated groups ($p < 0.01$)

4.6. Discussion

The results of our study provide compelling evidence of the effectiveness of RescuAR in self-directed CPR teaching and training. The significant improvements observed in both compression depth and frequency among participants in the experimental group demonstrate the positive impact of RescuAR on CPR skills. Moreover, the experimental group exhibited significantly greater improvements in effective CPR performance compared to the control group, underscoring the comprehensive benefits of RescuAR's AR-based approach. These findings suggest that traditional CPR teaching methods may have limi-

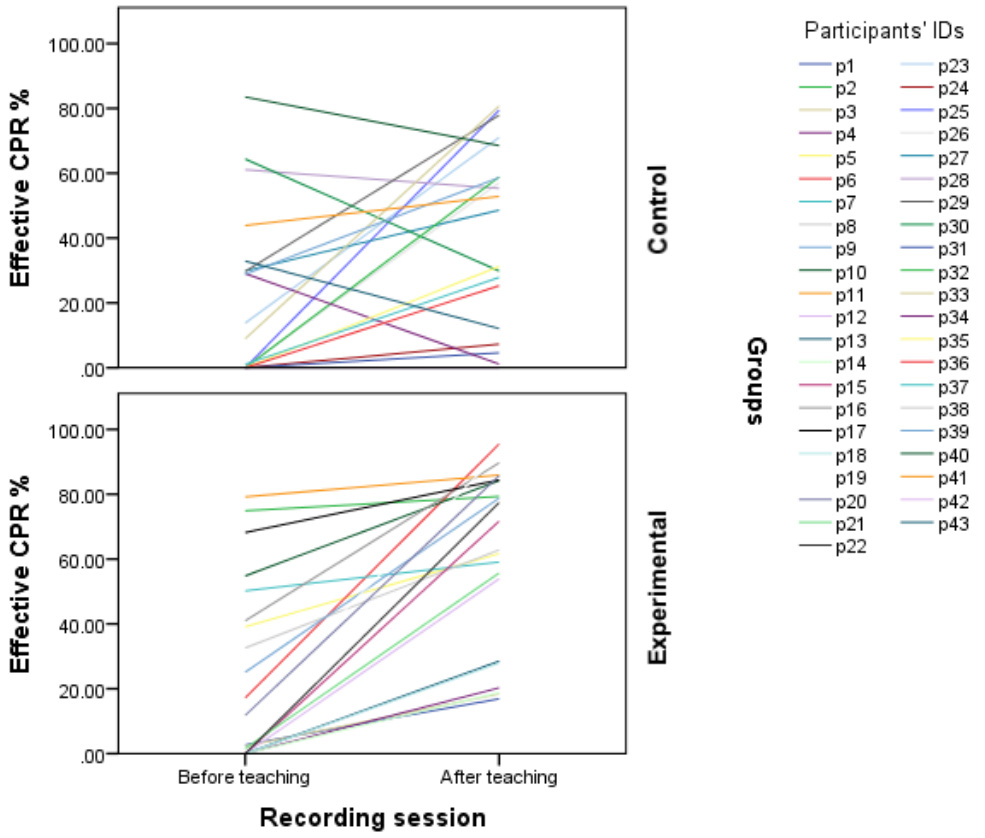


Figure 4.7.: Participant's performance lines before and after teaching sessions

tations in adequately addressing crucial aspects of skill acquisition and retention.

Acquiring physical skills like CPR relies on developing accurate muscle memory, and timely intervention is crucial to prevent the formation of incorrect habits. However, accurately assessing learner performance is challenging for teachers without measurement devices. Even with measurement systems, the timing and manner of intervention can impact learner performance. Previous research has explored the use of measurement systems to provide insights for teachers, but their effectiveness is influenced by various factors, including intervention timing and the method used [98, 278].

Our system showed that using real-time feedback modalities integrated into the teaching routine could help to overcome this challenge. These findings align with previous research highlighting the benefits of audio-visual feedback in CPR education [98, 140]. Furthermore, our approach for real-time detection and visualization of CC depth and frequency helped to seamlessly integrate the real-time feedback into the training routine, overcoming the limitations caused by using commercial CPR feedback devices, such as hand injuries due to the placement of the device, or instability of using smart devices for measurement during CC [8, 111].

Moreover, successful teaching requires sufficient dedicated time to achieve specific skills or abilities. However, in a traditional setting, time constraints and variations in individual capabilities make it challenging to provide adequate attention to each student. In a crowded classroom setup, many students may hesitate to ask questions or request repetitions during training sessions. This can hinder their learning experience and limit their understanding of the subject matter. **A self-directed approach — as implemented in RescuAR — promotes active participation, encourages question asking, and allows students to engage with the material at their own pace**.

Additionally, instructors themselves differ in their approaches and abilities to teach and evaluate effectively [132]. Consequently, evaluations or presentations dependent on human teachers may lack objectivity and standardization. To address these challenges, previous studies have attempted to unify teaching routines and enhance training sessions using various technologies. Some studies focused on teaching CPR principles through videos [29, 203, 268]. While unifying teaching is essential, learners' active engagement also plays a significant role in improving outcomes [232]. For instance, de Sena *et al.* found that although video-based teaching improved CPR skills, participants preferred more interactive and engaging self-training over passive video-based instruction [232]. Our study demonstrates that merging approaches can provide an efficient alternative for CPR education. **By utilizing a virtual teacher with programmed curricula, we standardized teaching routines without diminishing learners' interest, promoting verbal communication between teacher and student**.

In line with the findings of Balian *et al.* [17], our study further substantiates the potential of AR in CPR training. By incorporating a control group and measuring baseline performance, we provide additional evidence to support the effectiveness of AR for CPR training. This strengthens the understanding of the benefits and possibilities that AR technology offers in enhancing CPR education through various designs and implementations.

Although Leary [149] did not find a statistically significant difference between CPRReality training and a standard audio-visual feedback manikin, our study demonstrates significant improvements in CPR performance within the experimental group, suggesting the advantages of our AR-based approach in enhancing CPR quality compared to traditional methods.

4.6.1. Limitations

While our system design and study results provide valuable insights, it is essential to acknowledge their limitations. Certain technical issues, such as loss of environment mapping, occasional teacher avatar misplacement, and limited field of view, were identified during the experiment, indicating the need for further optimization. Additionally, the study's controlled environment and participants' awareness of evaluation may have influenced their performance, necessitating further validation of the app's real-world applicability in diverse settings with participants unaware of being evaluated. Furthermore, further research is needed to assess the long-term durability and retention of the im-

proved CPR performance observed in our study. Lastly, our study primarily focused on compression depth and frequency as key performance indicators; future studies could consider incorporating a more comprehensive assessment, including other critical elements such as hand placement, posture, and breath technique.

4.7. Conclusion

In conclusion, our study provides compelling evidence of the effectiveness of RescuAR, an AR-based CPR teaching and training app, in improving compression depth, frequency, and overall CPR performance. The insights gained from our study inform the development of guidelines and best practices for incorporating AR technology in CPR training programs. Policymakers, educators, and healthcare professionals can use this information to establish standards and recommendations for the implementation and utilization of AR-based training tools. This includes considerations such as the design of instructional content and feedback mechanisms, and the integration of AR training into existing curriculum frameworks. While this study primarily focused on the fundamental teaching of CPR, our future work will delve into exploring the integration of more realistic and immersive simulation scenarios. By incorporating advanced training systems, we aim to investigate ways of providing learners with a comprehensive and realistic training experience that prepares them for a wider range of CPR situations and challenges. By doing so we aim to contribute to expanding the field of research on medical training and inform the development of more effective CPR training programs that enhance learner performance, confidence, and ultimately save more lives.

Chapter 5

Decision-Making and Cognitive Support During High-Risk Procedures

Contents

AR technology holds significant promise not only in skill acquisition but also as a supportive and assistive tool in clinical practice. One of the most impactful applications of AR lies in providing visual guidance during cognitively challenging and demanding tasks. To this end, this chapter explores the use of AR as a cognitive support system during complex and high-stakes surgical procedures. By shifting the focus from educational settings to the more challenging, complex, sensitive, and time-critical environment of surgery, this work aims to deepen our understanding of the capabilities and limitations of current AR technologies and how they might be effectively integrated into clinical workflows.

This chapter presents the design, development, and system evaluation of ARAS, an AR assistance system for pancreatic tumor resection, one of the most complex and high-risk surgical procedures. The evaluations were conducted beyond the laboratory setting during a clinical trial involving 20 patients. Pancreatic surgery was specifically chosen due to the significant gap in the literature, largely driven by the extreme challenges associated with applying AR systems in this context. The real-world impact of ARAS, including surgical performance and patient outcomes, will be discussed in Chapter 6.

The content of this chapter covers the following one journal and three conference publications:

1. Javaheri, Hamraz, Omid Ghamarnejad, Ragnar Bade, Paul Lukowicz, Jakob Karolus, and Gregor Alexander Stavrou. "Beyond the visible: preliminary evaluation of the first wearable augmented reality assistance system for pancreatic surgery." International Journal of Computer Assisted Radiology and Surgery 20, no. 1 (2025): 117-129.

2. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. "Design and Clinical Evaluation of ARAS: An Augmented Reality Assistance System for Open Pancreatic Surgery." In 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 376-385. IEEE, 2024.

3. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, and Jakob Karolus. "From Concept to Clinic: Multidisciplinary Design, Development, and Clinical Validation of Augmented Reality-Assisted Open Pancreatic Surgery." In Proceed-

ings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1-24. 2025.

4. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. "ARAS: LLM-Supported Augmented Reality Assistance System for Pancreatic Surgery." In Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 176-180. 2024.

This work was awarded with "Best Demo Award" during ISWC/UbiComp'24 conference in Melbourne, Australia, where the full system and its capability was demonstrated as a demo paper.

5.1. Background

In theory, AR enables spatial visualization and interaction with life-sized 3D models, potentially allowing in-situ visualization of a patient's anatomical structures to aid surgical procedures. The primary advantage of AR lies in its ability to help surgeons locate vital structures, define dissection boundaries, and assist in surgical removal while minimizing the risk of damaging non-visible vessels, particularly in procedures like abdominal surgery that require precise dissection to avoid harm to surrounding tissues and vital vessels [81].

However, the efforts to develop AR-based surgical navigation and assistance systems for open abdominal surgery have been limited due to significant challenges and constraints. These include registration issues related to organ shifting and deformations [62, 195, 197], as well as environmental and performance constraints such as time pressure, sanitation requirements, high demands for system accuracy and robustness, and tracking difficulties caused by occlusion and lighting. These constraints often exceed the capabilities of lab-tested prototypes.

Only a few studies have been conducted in ecologically valid environments and tested in vivo during open surgery [25, 81]. Moreover, previous attempts at AR surgical assistance in abdominal surgery have largely been confined to tablet-based, handheld, or monitor-based solutions. While these approaches have provided valuable insights, they often fail to fully leverage the potential of modern wearable AR technology [168, 195, 200, 252], as they require surgeons to shift their focus away from the operating field to view information on separate displays, thereby disrupting the workflow [262].

In this work, we aim to address this gap by focusing on one of the most critical abdominal surgeries, tumor removal pancreatic surgery (PS) which stands out as an ideal candidate for AR-based navigation systems among abdominal surgeries for two key reasons. First, the high complexity of PS arises from the pancreas's unique location and its close proximity to major blood vessels, which makes the dissection process particularly challenging and risky [81]. This underscores the importance of navigational assistance. In this context, wearable AR can provide precise visualization of patient-specific 3D recon-

structed structures anchored to the patient's body, enabling safer dissection. By overlaying relevant information directly onto the surgical field in real-time, wearable AR technology can enhance situational awareness and improve the identification and localization of vital anatomical structures during surgery, helping to avoid accidental damage. Second, unlike other abdominal organs, which face registration issues related to organ shifting and deformations [62, 195, 197], the pancreas experiences minimal shifts and deformations throughout the surgical procedure due to its position at the back of the abdomen behind the peritoneum[1] [81]. This stability alleviates some of the constraints for open surgery and additionally makes it an ideal candidate to benefit from preoperatively captured data sources used for intraoperative navigation via in-situ visualization. Despite this advantage, there is still a notable gap in the literature regarding comprehensive interdisciplinary research on AR-based assistance and navigation systems in this domain [81, 280].

5.2. Methodology

Throughout this study, we employed a user-centered research methodology, involving end-users in design, development, and evaluation stages of ARAS. This process included the pre-design, design and development, lab testing, and refinement phases, followed by a preliminary clinical evaluation and an iterative cycle of clinical trials, further refinements, and additional testing as needed (Figure 5.1). This was done in close interdisciplinary collaboration between surgeons and the research team. To guide our design, development, and evaluation process, we addressed the following research questions:

Research Questions (RQs):

1. **What are the areas to improve with AR in the surgical domain throughout the perioperative process of pancreatic surgery, from surgeons' perspectives?**

   Objective: To identify, categorize, and evaluate the potential applications of AR technology in pancreatic surgery from the perspective of practicing surgeons, with a specific focus on its utility across the perioperative continuum, including preoperative planning, intraoperative guidance, and postoperative analysis to inform the development of clinically relevant and user-centered AR systems.

   Method In-depth interviews with experienced and novice surgeons from the field.

2. **How can we design an AR assistance system (ARAS) that seamlessly integrates into surgical workflows and aids in localizing vital structures in open PS?**

[1] The tissue that lines the abdominal wall and covers most of the organs in the abdomen.

a) ***User Requirements:*** What are surgeons' specific needs and preferences regarding the AR system's interface and functionality during open PS?

Objective: To gather insights from surgeons to inform the design of ARAS, ensuring it meets their practical requirements and enhances their ability to perform the surgery.

Method: Custom questionnaire on the importance of each surgical phase, interviews with surgeons, along with demographics of participating surgeons.

b) ***Technical Feasibility:*** What are the technical requirements and challenges for integrating anatomical data into an AR system used in open PS?

Objective: To identify constraints as well as feasible data, registration, and tracking methods for achieving in-situ augmented data visualization, ensuring that the AR system integrates seamlessly into the operating room and maintains the overall workflow without causing disruptions.

Method: Observations during surgical procedures and preliminary experiments on different viable approaches for registration and tracking.

3. **What is the intraoperative performance of ARAS in registering patient-specific 3D anatomical models during open PS?**

a) ***Accuracy:*** What is the intraoperative registration accuracy for the virtual anatomical models?

Objective: To determine how closely these 3D models reconstructed from preoperative data align with actual anatomical structures of the patient during the surgery.

Method: Postoperative questionnaire on registration accuracy completed by surgeons, focusing on specific structure groups including arteries, veins, and tumors.

b) ***Robustness:*** How well do the registered 3D models maintain their accuracy throughout different phases of the surgical procedure and across different types of open PSs?

Objective: To evaluate the impacts of anatomical deformation and surgical manipulations on the accuracy of the model registration.

Method: Postoperative questionnaire on registration accuracy of specific structure groups during each phase of surgery, including preresection, resection, and postresection across

different types of pancreatic tumor removal surgeries, including distal pancreatectomy, pancreaticoduodenectomy, and total pancreatectomy.

4. **How do surgeons perceive the usefulness of ARAS in enhancing anatomical localization during open PS?**

   Objective: To evaluate the surgeons' user experiences when working with ARAS for anatomical localization during open PS, including their perceptions of its effectiveness, its ease of use, as well as the overall impact on their surgical decision-making and workflow.

   Method: Intraoperative observations, postoperative interviews, post-trial UMUX-Lite questionnaire.

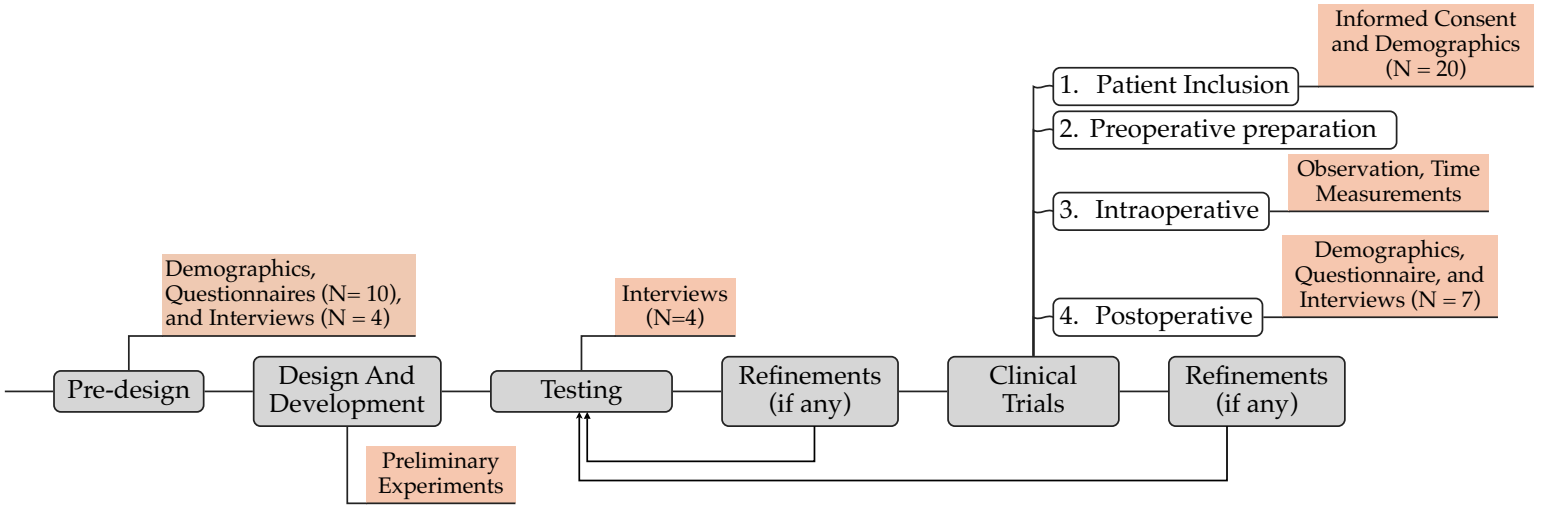


Figure 5.1.: Procedure chart showing different stages of the study from pre-design till the end of clinical trials and associated data collection in each phase.

5.2.1. Ethics

The study protocol received approval from the Ethics Committee of the Medical Association of Saarland under the registration number: 159/23. The protocol of our study was registered at ClinicalTrials.gov under the registration number: NCT06208579.

5.2.2. Materials

We utilized the Unity 3D game engine [258] for design and development of all software components of ARAS that were used in this study. We designed ARAS compatible with HoloLens 2 [180] and used this device as a wearable and head-mounted AR display throughout the study. The decision to use HoloLens 2 was motivated by our study results on better suitability of MH2 and see-through headset for critical and precision required tasks (see Chapter 3) and also mainly due to ethical, and safety considerations as it is CE-certified and was shown to have been successfully used in the medical domain before [121, 123, 160, 225]. Furthermore, 3D markers (Section 5.3.2) were designed in Bambu Studio [145] and subsequently 3D printed using the Bambu Lab P1S [144] for use in this study.

5.2.3. Data Collection

In the pre-design stage, we collected data using a custom pre-design questionnaire on factors such as criticality, time pressure, likelihood of complications, and the necessity of an assistance tool at each stage of open PS (preresection[2], resection[3], and postresection[4]). We also conducted in-depth semi-structured interviews with surgeons. Additionally, we collected the demographics of the participating surgeons.

During the design and development stage, we performed preliminary experiments targeting various functionalities of the system and conducted testing sessions with surgeons in a lab environment. Each testing session concluded with interviews to identify the required refinements. Each refinement session was followed by another testing session to validate the implemented system changes.

[2]The preresection phase encompasses the initial phase of the surgery spanning from the first big incision performed in the patient's abdominal area until the start of the resection phase. This phase involves preparing the blood vessels by carefully removing tissues attached to them, identifying target blood vessels either infiltrated by the tumor or essential to the procedure, and establishing strategies to minimize damage to surrounding healthy blood vessels and tissues before the resection begins.

[3]The resection phase covers the stages involving excision of the tumor and surrounding tissue, including the careful removal of any affected organs or blood vessels, while preserving as much healthy tissue as possible and ensuring clear margins to reduce the risk of tumor recurrence.

[4]Postresection phase covers the stages after resection until wound closure involving the closure of incisions, monitoring for complications such as bleeding and finally closing the patient's abdomen area.

After confirming that no further refinements were needed in the lab setting, a preliminary clinical evaluation phase was conducted. This phase involved testing the proposed approach and system in five patient cases. It was essential for validating the system's reliability and uncovering unforeseen challenges that could only emerge during real surgical procedures, challenges that could not be fully simulated in a laboratory environment. Following each surgery, in-depth interviews were conducted, leading to iterative system refinements and the implementation of new approaches based on the issues and insights identified during this phase.

Finally, after ensuring that no further refinement was needed, clinical trials commenced. During the clinical trials, patients' demographics and consent were collected. Additional data, including surgical time, system usage time, demographics of the participating surgeons, intraoperative observations, postoperative questionnaire responses on system registration accuracy, and postoperative in-depth semi-structured interviews, were also gathered.

At the end of all clinical trials, surgeons who participated in at least three surgery sessions and used ARAS for a minimum of 180 minutes (to avoid first-time user bias) were asked to complete a UMUX-Lite questionnaire [157]. This complemented the collected interviews, serving as a rapid and reliable tool for assessing new healthcare technologies [28].

5.2.4. Participants

10 surgeons (average surgical experience = 14 y, SD = 6.6 y) completed our custom pre-design questionnaire, and four surgeons (average surgical experience = 16.25 y, SD = 8.41 y) were interviewed during the pre-design stage. The same team of four surgeons participated in preliminary clinical evaluations.

During the clinical trials (Section 5.4), each surgical session was conducted by a team of four surgeons. In total, seven different surgeons (average surgical experience = 16 y, SD = 8.14 y) participated in the clinical trials specializing in Hepatobancreatobiliary [5] and Visceral [6] surgeries. Among these surgeons, four (average surgical experience = 14 y, SD = 6.82 y) used ARAS for more than 180 minutes during clinical trials and participated in at least three surgery sessions. These surgeons completed the UMUX-Lite questionnaire. The details of the surgeons' profiles, their roles in clinical trials, and their participation in various phases of data collection are provided in Table 5.1.

[5] Hepatopancreatobiliary surgery consists of the general surgical treatment for benign and malignant diseases of the liver, pancreas, gallbladder, and bile ducts.

[6] Visceral surgery, also known as abdominal surgery, refers to surgery of the abdominal cavity and abdominal wall, endocrine glands and soft tissue.

Table 5.1.: Profiles of participating surgeons throughout the study, their roles in clinical trials, and collected data; PDQ: pre-design questionnaire, PDI: pre-design interviews, PTI: post-testing interview, PORAQ: postoperative registration accuracy questionnaire, POI: postoperative interview, PTULQ: post clinical trials UMUX-Lite questionnaire.

ID	Surgical field	Surgical experience (years)	Surgical roles in clinical trials	Collected data					
				PDQ	PDI	PTI	PORAQ	POI	PTULQ
S1	Hepatopancreatobiliary	23	Lead, second lead	✓	✓	✓	✓	✓	✓
S2	Hepatopancreatobiliary	25	Second lead		✓		✓	✓	
S3	Hepatopancreatobiliary	13	Lead, second lead	✓	✓	✓	✓	✓	✓
S4	Hepatopancreatobiliary	16	Lead, second lead	✓			✓	✓	✓
S5	Visceral	25	Second lead	✓		✓	✓	✓	
S6	Visceral	4	Assistant	✓	✓	✓	✓	✓	✓
S7	Visceral	6	Assistant	✓			✓	✓	
S8	Visceral	9	None	✓					
S9	Visceral	17	None	✓					
S10	Visceral	18	None	✓					
S11	Visceral	9	None	✓					

5.2.5. Data Analysis

We used the R programming language [217] for data visualization and statistical analysis of our quantitative data, including the pre-design questionnaire, registration accuracy, UMUX-Lite, and operation-related metrics such as surgical time and system usage time.

All interviews throughout this project were transcribed verbatim. Given the volume of the data, we followed the pragmatic approach to qualitative analysis as recommended by Blandford *et al.* [26]. Initially, two researchers analyzed 25% of the data. Following this, we created a preliminary coding framework through iterative discussions using ATlast.ti [14]. The remainder of the interview data was then divided equally among the two researchers for coding. In a concluding discussion, the coding framework was further refined, and the final main themes were identified.

For observational data, the research team used observation sheets to record observational data. For the empirical analysis of the observations made during the various sessions, the affinity diagram method was applied [49]. Initially, all key observations were written on sticky notes. These notes were then organized by similar comments or thematic connections. Each note was reviewed, and through a consensus process, they were matched with related notes. This grouping continued until all the notes were categorized into appropriate clusters. Finally, relevant headings were assigned to these grouped notes, and discussions were conducted to explore the relationships between these headings.

5.3. ARAS System Design

We initiated the design process of ARAS by collecting a custom questionnaire from 10 surgeons (surgical experience = 14y, SD = 6.64) and conducting in-depth semi-structured interviews with four surgeons from the Hepato-Pancreato-Biliary and visceral surgery domain, as detailed in Table 5.1.

During the pre-design in-depth interviews, we discussed the design, limitations, and requirements of an AR-based assistance tool that could be used throughout the surgical therapy pipeline in pancreatic surgery, investigating the areas that could be improved with AR assistance.

The pre-design questionnaire, however, focused primarily on the intraoperative phase and gathered surgeons' subjective opinions regarding the criticality, time pressure, likelihood of complications, and the necessity of an assistance tool for each phase of the operation. This helped us better understand the significance of each surgical phase and guided our attention toward the areas that surgeons identified as the most challenging.

Based on preliminary interviews conducted during the pre-design phase, two major areas were identified where the integration of AR could significantly benefit surgeons: the preoperative and intraoperative phases of surgery. AR was considered highly valuable for enhancing surgical planning and decision-making prior to surgery, and it was also seen as a powerful tool for intraoper-

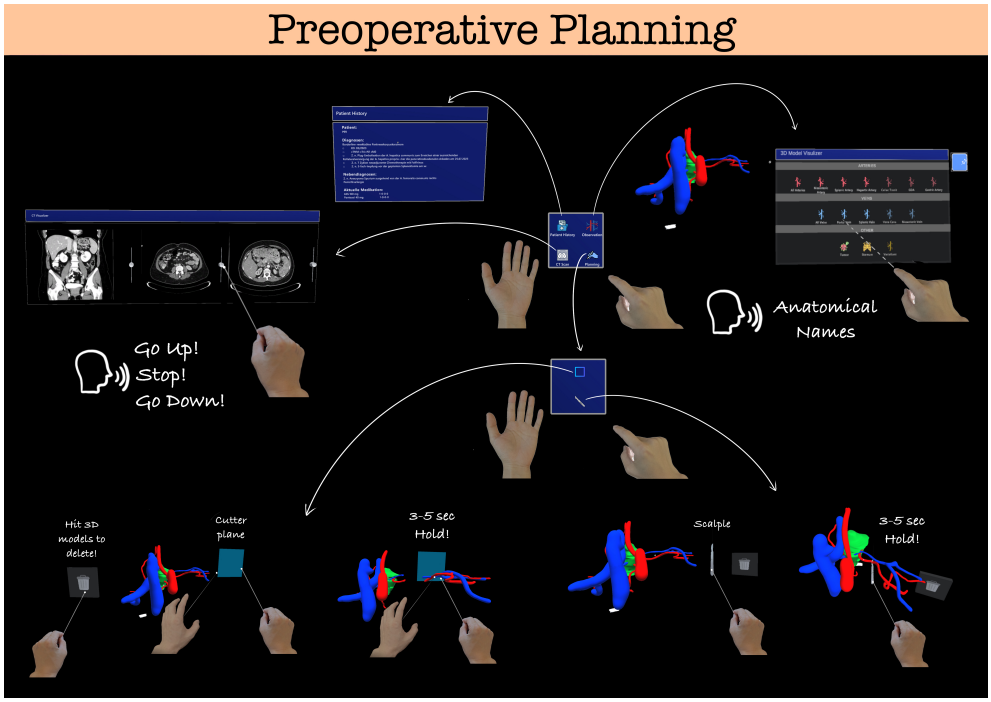


Figure 5.2.: Overview of ARAS planning mode features and implemented interaction modalities.

ative navigation.

In response to these findings, the ARAS main software was designed with two dedicated modes to support these distinct phases of pancreatic surgery. The preoperative mode facilitated detailed surgical planning, while the intraoperative mode served as a real-time navigation system to assist surgeons during the procedure. Additionally, a companion desktop application was designed to securely transfer patient data to the ARAS main application.

5.3.1. Preoperative Planning Mode

We designed ARAS-Planning to assist surgeons during the surgical planning session. In our design, we focused on implementing the features and functionalities essential for the surgical planning session, while also taking into consideration the users' needs and the limitations regarding interface and interaction methods.

Features and Functionalities

The system was designed to benefit surgeons in two primary ways. First, it provided comprehensive data visualization by integrating all relevant patient information, including medical history, CT slices, and a life-sized yet scalable 3D reconstruction of the patient's anatomy. This helped enhance spatial awareness and supported effective surgical planning. Second, it featured a resection simulation tool that allowed surgeons to perform virtual resections on

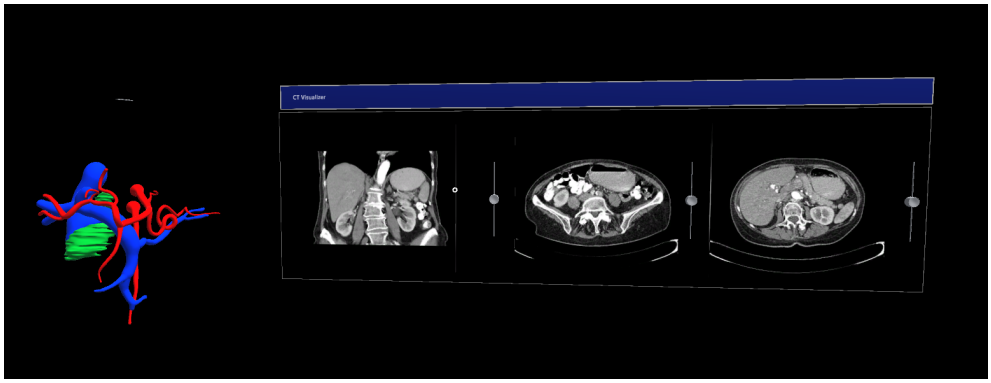


Figure 5.3.: Data visualization features implemented in ARAS-Planning

the 3D model and receive quantitative feedback, such as the estimated volume of remaining organ tissue. The surgeons could access all of these functionalities using a virtual menu attached to one of their hands. Together, these functionalities aimed to deliver a comprehensive planning tool for surgeons ahead of the intraoperative phase. An overview of system functionalities and features of ARAS planning mode is presented in Figure 5.2.

Data Visualization: One of the fundamental requirements specified by surgeons during the in-depth interviews was the capability to visualize the patient's vascular system and tumor in a 3D model of real-life size, and also the ability to further observe them in detail. This is crucial as sometimes making the connection between 2D CT images and the patient's anatomy can be challenging, especially for junior surgeons. Accordingly, we developed a feature that enables surgeons to visualize the patient's anatomical structures in 3D, alongside different slices of the CT images in the same view (Figure 5.3). Furthermore, we have introduced features that allow for the manipulation of the 3D model in 3D coordinates and the ability to scroll through the CT images as needed.

Resection Simulation: We have designed a resection simulation within ARAS-planning that allows surgeons to perform resections on the 3D reconstructed model prior to surgery, aiding them in developing a surgical plan. For this, we developed a dynamic mesh cutting approach with a scalable cutting plane. Surgeons could either use a manually scalable rectangular plane or use a dynamically scaling cutting plane attached to a virtual scalpel (see Figure 5.2). Surgeons can use either of these cutting planes to cut the target 3D structure at the desired point and angle, with the resection occurring upon 3 seconds of collision between the plane and the target structure (Figure 5.4).

To prevent accidental cuts to nearby 3D structures using a scalpel, we employed an adaptive scaling method. A non-visible cutting plane was initially attached to the tip of the virtual scalpel. Our method dynamically adjusts the width of this cutting plane in Unity based on the cross-sectional diameter of a 3D object to cut. Using HoloLens 2 eye tracking, first, a ray is cast from

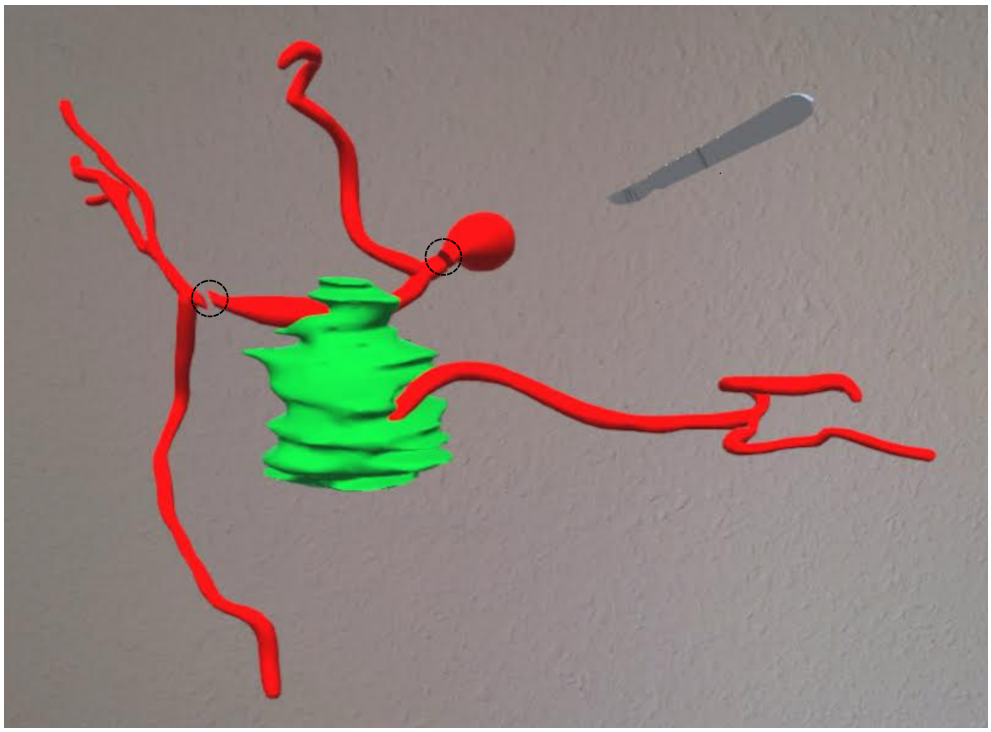


Figure 5.4.: Resection simulation implemented in ARAS-Planning

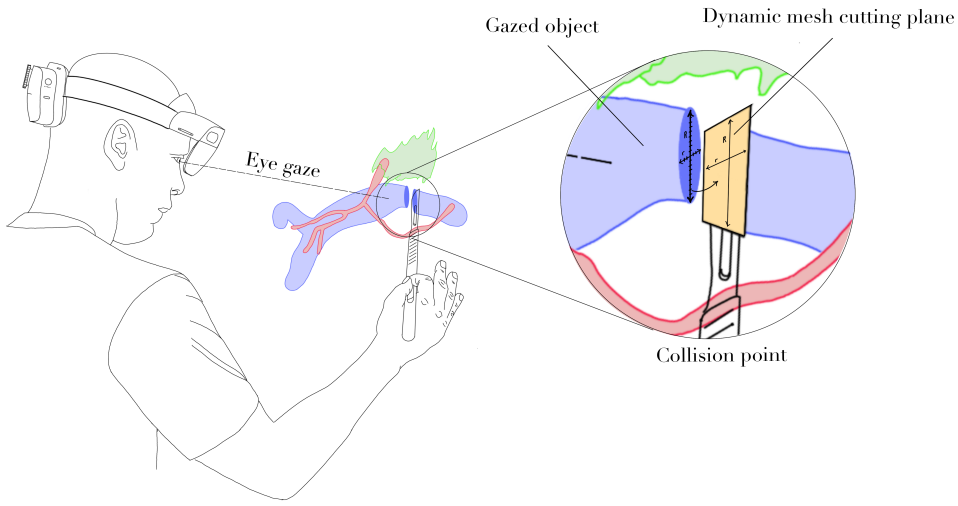


Figure 5.5.: Dynamic resection plane scaling with the radius of the gazed object at the collision point

the user's eye in the direction of their gaze. This raycast allows the system to identify both the object being observed. This raycast returns the list of hit meshes. The system constructs a temporary plane at the first hit location to the user (considering only the nearest gazed object), facing in the same direction as the scalpel. It then calculates where the surface mesh triangles of the object to cut (e.g., 3D model of a vessel) intersect with this temporary plane by checking each edge of a triangle to see if it crosses the cutting plane. If it does, the point of intersection is computed via linear interpolation and stored. Then those intersection points are projected into a 2D plane-aligned space, and the maximum distance between them represents the local cross-sectional diameter. This diameter is used to scale the cutting plane's width (attached to the scalpel), ensuring it matches the size of the object being examined (Figure 5.5). The c# code of the algorithm is given in Listing 5.1.

```
1     public Transform cuttingPlane;
2     public float gazeMaxDistance = 10.0f;
3
4     private IMixedRealityEyeGazeProvider eyeGazeProvider;
5
6     void Start()
7     {
8         eyeGazeProvider = CoreServices.InputSystem?.
              EyeGazeProvider;
9         if (eyeGazeProvider == null)
10        {
11            Debug.LogError(''Eye Gaze Provider not available.'');
12        }
13    }
14
15    void Update()
16    {
17        if (eyeGazeProvider == null || !eyeGazeProvider.
              IsEyeGazeValid)
18            return;
19
20        Ray gazeRay = new Ray(eyeGazeProvider.GazeOrigin,
              eyeGazeProvider.GazeDirection);
21        if (Physics.Raycast(gazeRay, out RaycastHit hit,
              gazeMaxDistance))
22        {
23            MeshCollider meshCollider = hit.collider as
                  MeshCollider;
24            Mesh mesh = meshCollider?.sharedMesh;
25            if (mesh == null) return;
26
27
28            Vector3[] worldVertices = mesh.vertices
29                .Select(v => meshCollider.transform.
                      TransformPoint(v))
30                .ToArray();
31
32            Plane cuttingPlaneWorld = new Plane(
33                cuttingPlane.forward,
34                hit.point
35            );
```

```
36
37
38              List<Vector3> intersectionPoints = new List<Vector3
                    >();
39              int[] triangles = mesh.triangles;
40
41              for (int i = 0; i < triangles.Length; i += 3)
42              {
43                  Vector3 v0 = worldVertices[triangles[i]];
44                  Vector3 v1 = worldVertices[triangles[i + 1]];
45                  Vector3 v2 = worldVertices[triangles[i + 2]];
46
47                  IntersectTriangleWithPlane(cuttingPlaneWorld, v0,
                        v1, v2, intersectionPoints);
48              }
49
50              if (intersectionPoints.Count >= 2)
51              {
52
53                  Vector3 planeNormal = cuttingPlaneWorld.normal;
54                  Vector3 planeRight = Vector3.Cross(planeNormal,
                        Vector3.up).normalized;
55                  if (planeRight.magnitude < 0.01f) planeRight =
                        Vector3.Cross(planeNormal, Vector3.forward).
                        normalized;
56                  Vector3 planeUp = Vector3.Cross(planeNormal,
                        planeRight);
57
58                  List<Vector2> projectedPoints =
                        intersectionPoints
59                      .Select(p => new Vector2(Vector3.Dot(p - hit.
                            point, planeRight), Vector3.Dot(p - hit.
                            point, planeUp)))
60                      .ToList();
61
62                  float maxDiameter = 0f;
63                  for (int i = 0; i < projectedPoints.Count; i++)
64                  {
65                      for (int j = i + 1; j < projectedPoints.Count
                            ; j++)
66                      {
67                          float dist = Vector2.Distance(
                                projectedPoints[i], projectedPoints[j
                                ]);
68                          if (dist > maxDiameter)
69                              maxDiameter = dist;
70                      }
71                  }
72
73                  (localScale.x assumed to be width axis)
74                  Vector3 newScale = cuttingPlane.localScale;
75                  newScale.x = maxDiameter;
76                  cuttingPlane.localScale = newScale;
77              }
78          }
79      }
80
```

```
81      void IntersectTriangleWithPlane(Plane plane, Vector3 v0,
            Vector3 v1, Vector3 v2, List<Vector3> result)
82      {
83          Vector3[] verts = new Vector3[] { v0, v1, v2 };
84          float[] distances = verts.Select(v => plane.
                GetDistanceToPoint(v)).ToArray();
85
86          for (int i = 0; i < 3; i++)
87          {
88              int j = (i + 1) % 3;
89
90              float d1 = distances[i];
91              float d2 = distances[j];
92
93              if ((d1 > 0 && d2 < 0) || (d1 < 0 && d2 > 0))
94              {
95                  float t = d1 / (d1 - d2);
96                  Vector3 point = Vector3.Lerp(verts[i], verts[j],
                        t);
97                  result.Add(point);
98              }
99          }
100     }
```

Listing 5.1: Cutting Plane Scaler Algorithm

A cut is made if the virtual cutting plane maintains contact with the target object to cut for at least 3 seconds, dividing the target 3D object mesh into two separate pieces at the collision point while maintaining the same material for the cut mesh plane surface.

For our mesh splitting simulating resections, we employed a mesh segmentation approach based on a flood fill algorithm to separate the affected mesh into distinct, connected components following a virtual cut. After performing the mesh slicing operation (3 seconds of contact between the cutting plane and object to cut), we used the topological relationships between triangle vertices to identify clusters of connected geometry. This was achieved by constructing a graph where each unique vertex stores references to its neighboring vertices, derived from the triangle connectivity. A flood fill traversal was then applied to label all connected vertices with a common cluster index. These clusters were subsequently used to generate individual mesh segments, allowing the simulation to accurately represent separate anatomical regions post-resection. The flow chart of segmentatiion algoirthm is given in Figure 5.6.

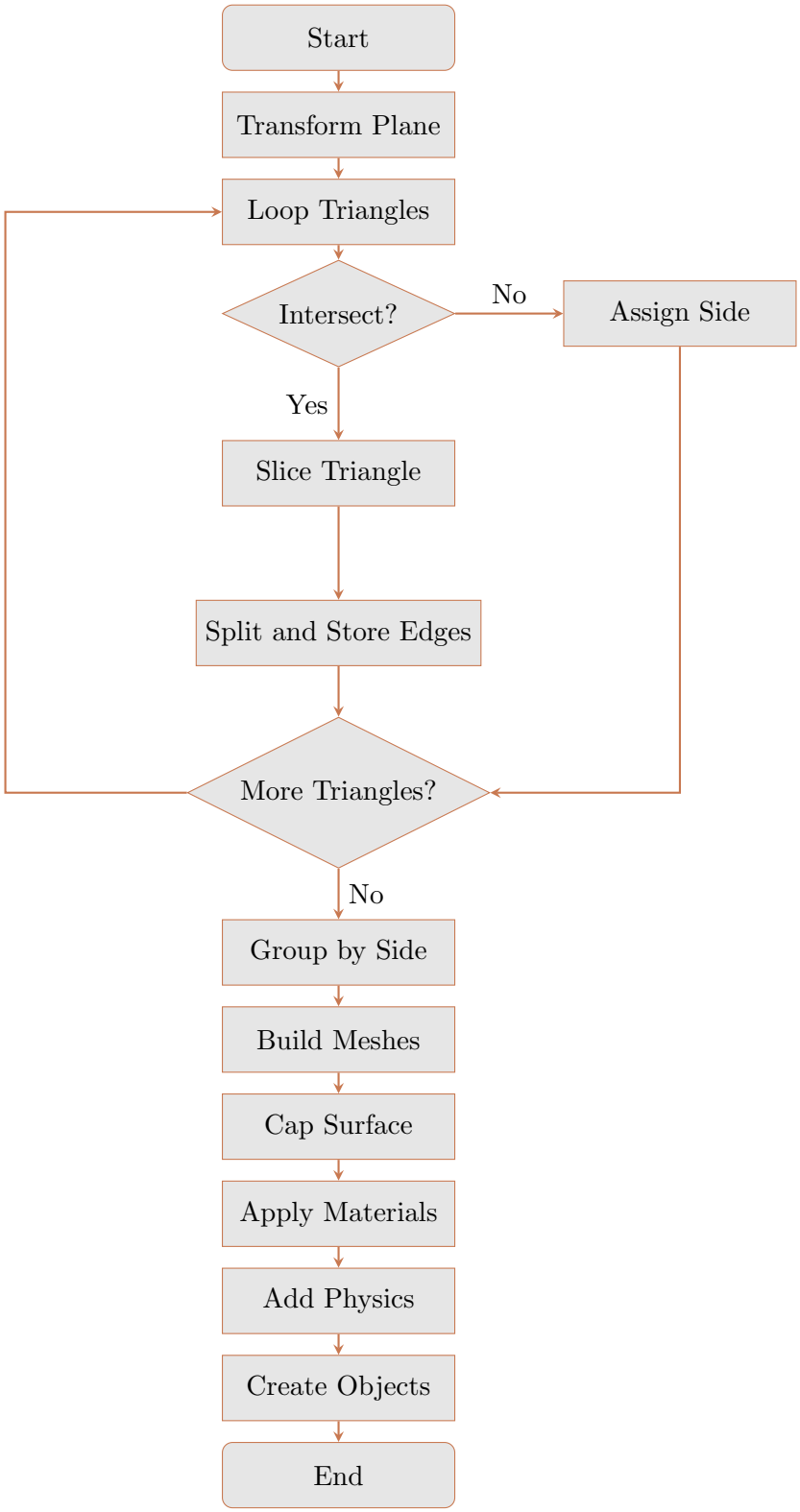


Figure 5.6.: Flowchart of the mesh slicing algorithm, including triangle classification, edge capping, and mesh reconstruction.

The code snippet presented in Listing 5.2 shows how the mesh is scanned triangle-by-triangle to build a vertex neighbor graph, where each vertex knows its directly connected neighbors. This forms the basis for identifying topological clusters.

```
foreach (KeyValuePair<Vector3, VertexInfo> entry in
    uniqueVertices)
{
    List<int> occassions = entry.Value.Occasions_Vertex;
    List<Vector3> neighbors = entry.Value.Neighbors;
    for (int j = 0; j < occassions.Count; j++)
    {
        int index = occassions[j];
        int r = index % 3;
        Vector3 b, c;

        switch (r)
        {
            case 0:
                b = vertices[index + 1];
                c = vertices[index + 2];
                break;
            case 1:
                b = vertices[index - 1];
                c = vertices[index + 1];
                break;
            case 2:
                b = vertices[index - 2];
                c = vertices[index - 1];
                break;
        }
        if (!neighbors.Contains(b)) neighbors.Add(b);
        if (!neighbors.Contains(c)) neighbors.Add(c);
    }
}
```

Listing 5.2: Vertex Adjacency Construction

The core flood fill loop expands from one unvisited vertex to its neighbors, recursively marking all connected vertices as part of the same cluster (Listing 5.3).

```
List<Vector3> unvisited = uniqueVertices.Keys.ToList();
List<Vector3> visited = new List<Vector3>();
List<Vector3> discoveryQueue = new List<Vector3>();

void AddUnvisitedNeighbors(Vector3 vertex)
{
    List<Vector3> neighbors = uniqueVertices[vertex].Neighbors;
    foreach (var neighbor in neighbors)
    {
        if (!visited.Contains(neighbor) && !discoveryQueue.
            Contains(neighbor))
            discoveryQueue.Add(neighbor);
    }
}

```

```
Vector3 currentVertex;
int group = 0;

while (unvisited.Count > 0)
{
    discoveryQueue.Add(unvisited[0]);

    while (discoveryQueue.Count > 0)
    {
        currentVertex = discoveryQueue[0];
        discoveryQueue.RemoveAt(0);
        unvisited.Remove(currentVertex);
        visited.Add(currentVertex);
        AddUnvisitedNeighbors(currentVertex);

        var info = uniqueVertices[currentVertex];
        for (int i = 0; i < info.Occasions_Vertex.Count; i++)
        {
            cluster[info.Occasions_Vertex[i]] = group;
            subMesh[info.Occasions_Vertex[i]] = info.
                Occasions_Submesh[i];
        }
    }

    group++;
    visited.Clear();
    discoveryQueue.Clear();
}
```

Listing 5.3: Flood fill to group connected vertices into clusters

Once the clusters are identified, this loop builds a separate mesh for each cluster by grouping triangles that share the same cluster ID. These become the new resection segments (Listing 5.4).

```
for (int i = 0; i < triangles.Count; i += 3)
{
    int[] tri = new int[3] {
        triangles[i],
        triangles[i + 1],
        triangles[i + 2]
    };

    int cluster = dynamicMesh.Cluster[tri[0]];
    int submesh = dynamicMesh.SubMesh[tri[0]];

    dynamicMeshes[cluster].AddTriangle(tri, submesh, dynamicMesh)
        ;
}
```

Listing 5.4: Separating triangles into per-cluster mesh segments

Upon creation of two mesh segments, the user could toggle behavior between automatic deletion of one of the resected parts or manually remove the part that is no longer needed. The automatic deletion algorithm relied on two rules to find the structure to delete.

1. If one of the newly created resection parts has no longer any collision points with the remaining of the 3D objects (other vessels or anatomical structures), that means it is completely disconnected, so it could be removed.

2. If there are more than 1 disconnected structure, it would calculate the center of the majority of the 3D structure and remove the part that remains furthest from the center.

However, users have the option to disable this automatic deletion and manually remove the resected parts by grabbing and disposing of them in a virtual trash.

User Interface and Interaction Methods

Since preoperative planning involves fewer constraints than intraoperative procedures in terms of environmental and user-dependent limitations, various interaction modalities were employed. We utilized hand tracking, eye tracking, and speech commands to facilitate interaction with the system. A virtual menu, attached to the palm of the user's hand via hand tracking, displays the main functionalities of the app, including CT images, patient history, 3D structure visualization, and resection simulation Figure 5.2. Users can interact with this hand menu by pressing virtual buttons with a finger or using speech commands designated for each button. Furthermore, we introduced a secondary virtual menu, accessible from the hand menu, which provides a list of interactable buttons for each reconstructed 3D structure. This allows users to enable or disable the visualization of each 3D structure.

Furthermore, we also assigned one or more unique keywords to each anatomical structure, facilitating the enabling or disabling of their visualization using speech commands (Table 5.2). In implementing these speech commands, we utilized the Windows speech recognition service (using IMixedRealityDictationSystem service) [271] and Microsoft Mixed Reality Toolkit, MRTK2 [116], to define specific keywords for triggering desired functionalities. These keywords were carefully selected in collaboration with surgeons from the field, ensuring relevance and ease of recall.

A total of 31 unique keywords were assigned to trigger the system functionalities (Table 5.2). Designed as medical terms commonly used in daily practice or intuitive words, each keyword corresponded either to the Latin or English names of anatomical structures available in the 3D model (Figure 5.9) or as intuitive words routinely used in a clinical environment or daily life. Recognizing the challenges of keyword recognition and aiming to enhance flexibility, some functions could be executed using multiple keywords or synonyms. For instance, the keyword "cancer" could be used interchangeably with the keyword "tumor" to enable or disable tumor visualization (Table 5.2). Additionally, all the keywords used for 3D visualization could be used in a combination of ON/OFF to either enable or disable some segments' visualization, or without these extensions as a toggle behavior between on and off mode. In addition

to speech commands specific to each 3D structure and menu buttons, we implemented control speech commands such as 'go up/down' and 'stop' to scroll through CT images. This feature addresses the challenges of hand interaction with objects out of reach, which can be difficult for those unaccustomed to interacting with wearable AR devices. To provide visual feedback to the user confirming the successful recognition of the spoken keyword, we used the MRTK tooltip component to be shown upon the detection of a speech command.

Table 5.2.: Implemented voice keywords and their functionalities

Type	**Voice Keywords**	**Functionality**
Patient information	Patient history / diagnosis / medication	Activates/Deactivates patient history panel
	CT / CT Image / Tomography / Computed Tomography / CT scans	Activates/Deactivates CT Visualizer
3D visualization	arteries/veins	Activates/Deactivates rendering of all arteries/veins
	Sternum	Activates/Deactivates rendering of the associated structure
	Celiac trunk	
	GDA	
	Mesenteric artery	
	Splenic artery	
	Gastric artery	
	Hepatic artery / liver artery	
	Portal vein	
	Vena cava	
	Splenic vein	
	Mesenteric vein	
	Tumor / lesion / cancer	
	Variation	Activates/Deactivates rendering of unusual anatomical structures
Control commands	Go up/down	Activates the automatic scrolling up/down of the CT slices
	Stop	Stops the automatic scrolling of CT slices
	Capture photo / hologram	capture a photo using HoloLens mounted camera with/without holograms

5.3.2. Intraoperative Assistance Mode

The development of the intraoperative mode of ARAS was initially guided by insights from pre-design questionnaires and interviews. The developed system, however, was iteratively refined using data collected during testing and clinical trial sessions.

The questionnaire on different phases of operation suggested that, while all stages of the operation are considered highly critical, the initial two stages of the operation (before tumor removal and resection phase) require more attention in terms of the need for an assistance tool as it has higher time pressure and the chance of intraoperative complication (Figure 5.7). Hence, while our system aimed to provide assistance throughout the operation, it prioritized the implementation of approaches targeting the initial two stages of the surgery, preresection and resection phases, which were characterized by the greatest need for AR assistance, as well as higher time pressure and criticality.

The ARAS intraoperative contained files on the patient's medical history and recent CT slices (taken within the last two weeks), as well as 3D models of the patient's vascular system and tumor. These 3D models are reconstructed from the CT scan slices by surgeons using the MeVis Liver Suites software [178] and were transferred to ARAS using a developed companion application. An overview of features and interaction modalities of ARAS intraoperative mode is given in Figure 5.8.

The final design of the ARAS software included two distinct feature sets to support surgeons throughout the surgical session: in-situ visualization and supportive data visualization. Due to differences in accuracy requirements, robustness, and the spatial and user-related constraints associated with each feature, they were designed with distinct interaction capabilities and approaches. For easy setup, after software initiation, the in-situ visualization mode would be automatically activated.

In-situ Visualization

The primary feature of the ARAS software was to provide navigational support specifically during the preresection and resection phases of surgery, following results of a pre-design questionnaire (Figure 5.7).

To achieve navigational in-situ visualization, we used a similar approach as previous works [252] and used patient-specific 3D reconstructed models of the vital vascular system and tumor, which are crucial for PS. For each patient, 12 segments were reconstructed from CT images. These segments included six arteries (celiac trunk, gastroduodenal artery, mesenteric artery, splenic artery, gastric artery, and hepatic artery), four veins (portal vein, vena cava, splenic vein, and superior and inferior veins), as well as the tumor and sternum. These arteries and veins, along with their location around the pancreas, are depicted in Figure 5.9. The 3D reconstruction process was performed by surgeons as part of their clinical routine for surgical planning, thus adding no additional workload for them.

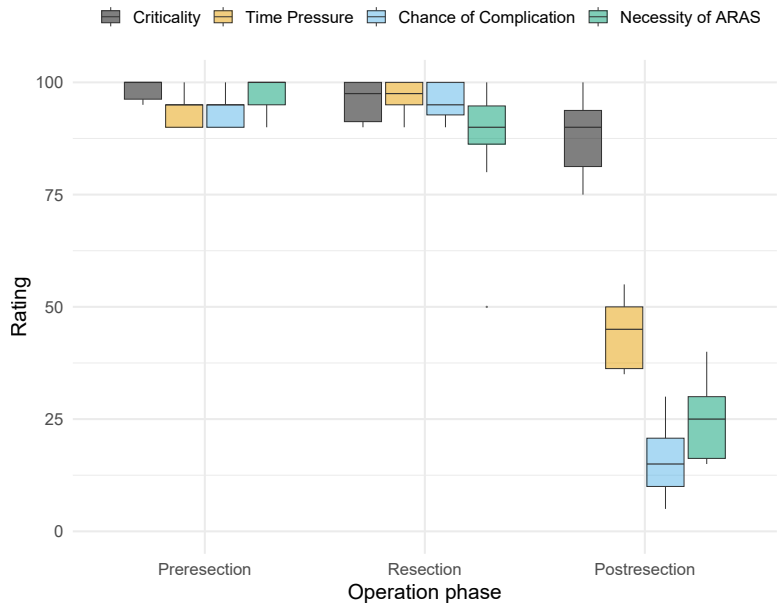


Figure 5.7.: Surgeons' subjective rating on criticality, time pressure, chance of intraoperative complication, and the necessity of having an assistance tool during each distinct surgical phase.

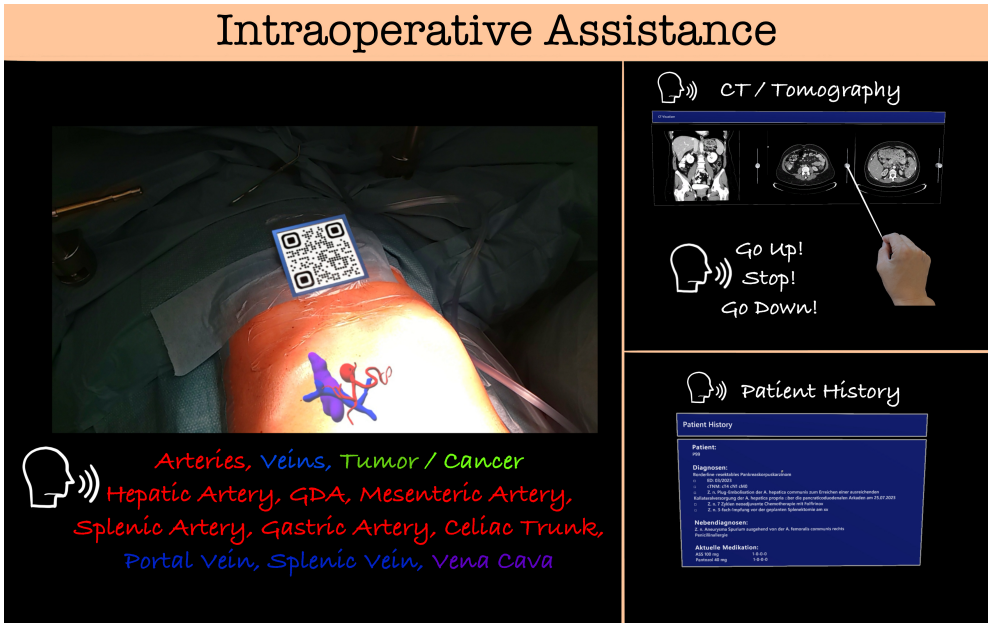


Figure 5.8.: Overview of ARAS Intraoperative mode features and implemented interaction modalities.

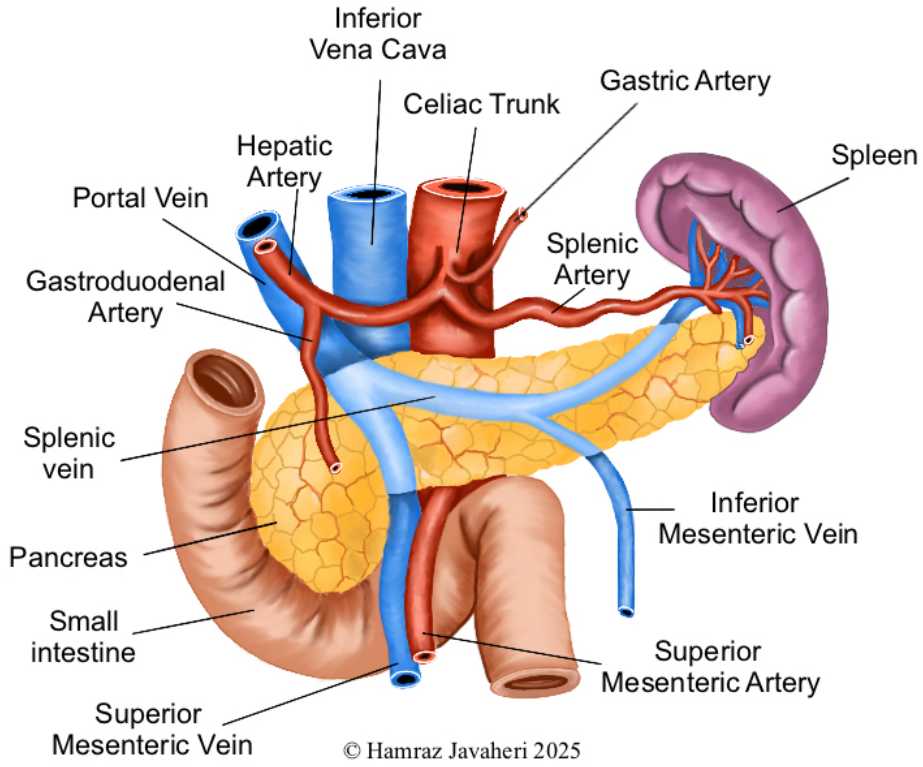


Figure 5.9.: Anatomical drawing of the vital veins and arteries that 3D reconstructed as segments and their positions around the pancreas.

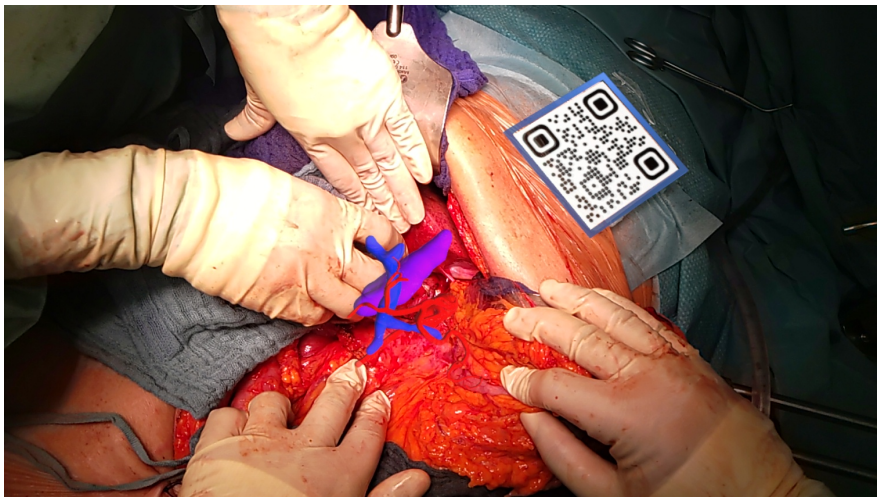
These reconstructed 3D models of patient were used to provide in-situ navigational guidance by overlaying them onto the patient's abdominal area. In Table 5.3, a snapshot captured by the HoloLens 2 is shown, in which the in-situ visualization of a patient's anatomical structures is presented, registered, and tracked using a marker during a surgical procedure. The characteristics of the features and the interaction modalities used for in-situ visualization are also provided in the same table.

Dual-Layer Marker-based Registration and Tracking To ensure an accurate overlay of the virtual patient-specific 3D model, we opted for a marker-based tracking approach similar to previous works [151, 252], using RGB cameras on the HoloLens headset and Vuforia SDK [213]. This choice addressed the challenges of constantly changing scenery and performance considerations. To achieve robust feature tracking and maintain accuracy, continuous tracking of traceable features was essential.

Many state-of-the-art registration methods lead to several issues, including disruptions during the surgical procedure, such as those caused by manual registration, point-based registration, and surface registration requiring initial alignment or additional need for hardware like a 3D scanner [164].

While alternative approaches relying on anatomical features [186] or various sensing modalities such as ultrasound [226] and depth sensors [238, 269, 273]

Table 5.3.: The characteristics of in-situ visualization and the feature set provided by the ARAS software.



In-situ Visualization

Data type	3D reconstructed vascular system and tumor
Placement method	Marker-based
Features	Segment-based visualization
	Enable/disable marker tracking
	Automatic re-orientation of 3D model on marker detection
	Freeze 3D model position and orientation
Interaction method	Keyword based voice control

may be effective for minimally invasive surgeries or surgeries involving non-deformable organs [242], our preliminary experimentation with such methods showed they often encounter difficulties with accurately registering the previously reconstructed 3D model onto organs due to organ deformation and the dynamic operation field, which includes the presence of the surgeon's hands, blood, lap sponges, and tools used during open abdominal surgery.

In contrast, marker-based tracking using RGB cameras offered several advantages. The artificial and predefined features of the markers remain unchanged and traceable throughout the procedure. This method allowed for indirect registration of the 3D models of the vascular system by placing markers outside the operating field. Furthermore, it enabled reliable tracking in a highly variable and restricted environment using unchanging artificial features, without the need for additional devices or computational power.

We performed a preliminary evaluation on the robustness and feasibility of different marker types and materials for the surgical environment. We have considered light intensity (lx/m^2) and the sterilization process as two main affecting factors during the marker design process to be used in a surgical environment. While the extreme change in light intensity throughout the operation might affect the reflection and traceability of the marker, the sterilization process might cause a loss in the integrity of the marker features. We have considered paper print and 3D print as two viable solutions to print the marker. While paper-printed and 3D printed markers both could be placed in a sterile plastic cover, only 3D printed markers could undergo a sterilization process. We tested PLA, ABS, and PETG filament types to print the 3D marker and evaluated the integrity of the marker features after a standard steam sterilization process. In line with *Dautzenberg*'s work [57], we have found out that PLA demonstrated fewer morphological deviations after sterilization, therefore preserving the integrity of marker features the most.

We used Vuforia [213] marker tracking with an enabled extended tracking method to track the marker. We designed a 7cm x 7cm image target with the highest traceability rating (5 out of 5) from Vuforia as our marker. We printed our designed marker both on paper and also 3D printed this marker using 3 different approaches (Figure 5.10). We used three combinations of mat white PLA background and mat black PLA features, fluorescent white PLA background and mat black feature, and mat black PLA background and fluorescent white PLA features for our 3D printed marker. We evaluated the traceability of each of these markers using the HoloLens 2 camera in a surgical operation room using surgical lights under 13 different light illuminances over ten cycles, each one minute. We placed a marker on a surgery table and measured the illuminance where the marker was placed. We chose a range of 50 lx (lux) (dimmed room) and a maximum of 120 klx (highest illuminance on surgical light) as our illuminance range. All 3D printed markers performed better than the plastic-covered paper-printed marker, especially under activated surgical lights. 3D printed marker with a fluorescent white background and black mat features demonstrated the highest robustness to different light conditions (Figure 5.10). Considering the sterilization and robustness to dif-

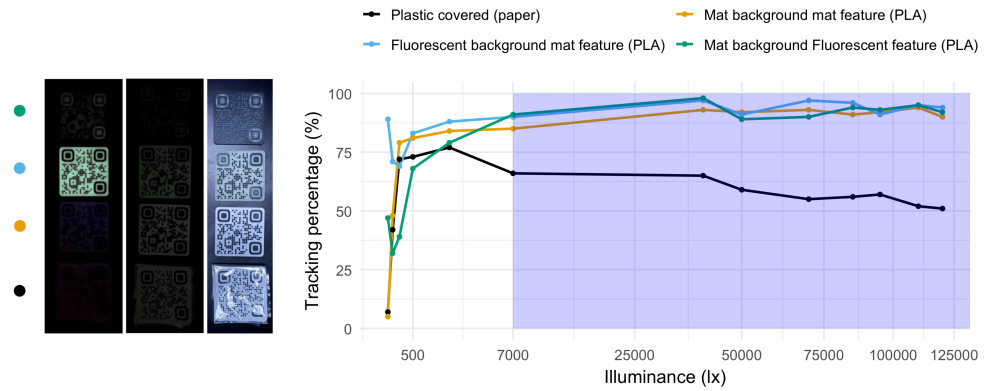


Figure 5.10.: Tracking percentage of different marker types under different illuminance conditions using Vuforia and HoloLens 2 camera. The blue region shows the illuminance interval achieved with different intensity levels of surgical light.

ferent light conditions, we hence used the designed marker with fluorescent white PLA background and matte black feature as our marker throughout our study.

For the selection of registration point and marker placement, we adopted a different approach to previous work demonstrated by Tang et al. [252], which involved direct placement of the marker on the target organ. Through pre-design investigations and detailed post-testing interviews with surgeons, we identified the optimal strategy for marker placement. This process led us to adopt an indirect approach, positioning the marker outside the operational field for improved effectiveness. While directly attaching markers to the target organ, such as the pancreas, or vascular system, could yield precise 3D model registration [252], it would require manual placement and removal of the marker in order to continue the surgical procedure. Therefore, we considered using a stable anchor point that remains fixed during surgery and maintains a consistent positional and rotational relationship with the target structures. Our research suggested that the xiphoid process of the sternum, located at the center of the chest just below the lower part of the sternum (Figure 5.11), was a promising anchor due to its stability throughout PS and also enabling active marker tracking to compensate for the patient's respiratory motion. Given the minimal mobilization of organs during this type of surgery, the positional and rotational relationships between the xiphoid process and the vital structures involved in PS are consistently maintained. Consequently, we chose the xiphoid process of the sternum as our anchor point for marker attachment.

For each patient, we measured the distance from the end of the xiphoid process to the jugular notch, as well as the distance from the xiphoid process to the patient's skin layer with millimeter accuracy using CT images through the MeVis software [178]. The distance to the jugular notch was used to ensure accurate marker placement, while the distance to the skin layer helped adjust the virtual object's position relative to the marker's surface. The marker was placed flat on the xiphoid process while ensuring the correct distance and

angle to the jugular notch so that the end of the marker would be aligned with the end of the xiphoid process (Figure 5.11 (A)). The sternum marker was continuously tracked, and the model position was updated accordingly (compensating patient's respiratory motion).

However, relying solely on tracking the sternum for model registration presented certain limitations, which were identified in our preliminary system evaluation, particularly when surgeons occasionally needed to mobilize the rib cage to create more space for the operation. In such scenarios, the previously measured position and orientation of the 3D model relative to the registration point on the sternum became invalid, leading to a loss of registration accuracy. To address this issue, we investigated a second reliable anchor point (dual-layer approach) to update the relative position of the 3D model to the sternum marker. Due to its stable orientation relative to other vascular structures and its location within the inferior abdominal muscles and tissues, we found that the inferior vena cava could be used as our secondary registration anchor for cases where the sternum is shifted or mobilized to update the changed relative position of the 3D model to the sternum marker. Unlike the sternum, which can be identified from the top layer of the skin tissue, the vena cava is located in the deeper layers of the abdomen as depicted in Figure 5.11 (B). Therefore, this anchor point can only be used after the mobilization of the head of the pancreas during the later stages of the preresection vascular preparation, as it is otherwise hidden behind the pancreas and small intestine.

To choose a precise location for this anchor point that could be identified by surgeons, we selected the position where the left renal vein branch appears (Figure 5.11 (B)). Similar to the sternum marker, we calculated the position and orientation of all reconstructed 3D segments relative to this second registration point using CT images.

Since placing any marker in this delicate position was not feasible and the marker feature would not be traceable within the abdominal cavity, we designed a specific legged marker with rounded tips (to avoid damaging the vein) and fixed leg length that could be used to indicate this precise point without the need to placing the marker (the part that contains traceable features) directly on the vena cava (Figure 5.11 (B)). The surgeons used this marker to point out the registration anchor point at the branching point of the left renal vein from the vena cava, aligning the virtual model with its real structure, then used a keyword to update the model's position and orientation relative to the sternum marker before removing the marker. Upon detection of the vena cava marker, the 3D model would automatically reorient itself using previously stored data from CT images. If the user updates the position and orientation of the 3D model, the initial orientation of the sternum marker is overwritten. To reset the default orientation of the 3D model to the sternum marker, an explicit user request is required. After updating the new orientation and position relative to the sternum marker using the vena cava marker, the sternum marker is used for tracking again and continuously updating the model position (compensating for the patient's respiratory motion).

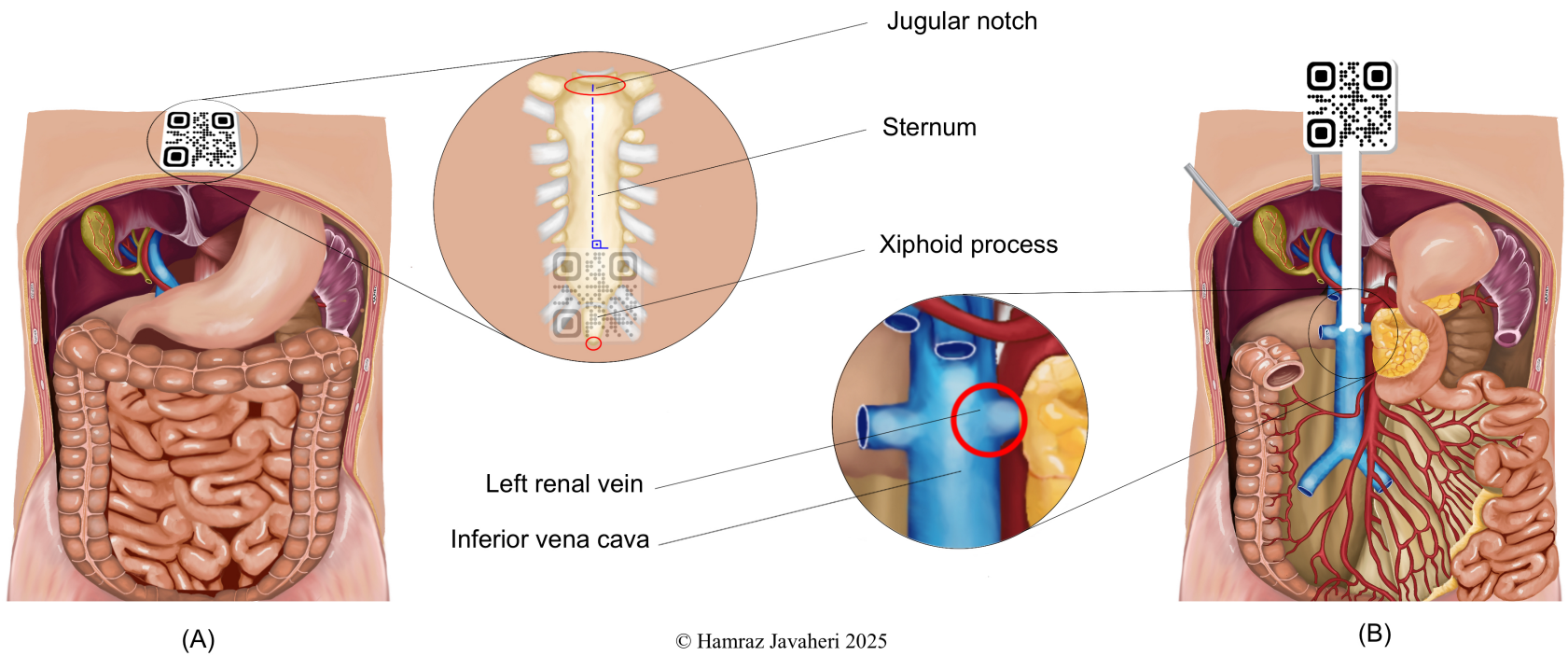


Figure 5.11.: Drawing depicting dual-layer marker placement on the sternum and vena cava. (A) The marker placement aligned with the end xiphoid process of the sternum. (B) The marker placement at later stages of the surgery, after vascular preparation and gaining access to the vena cava and registration point at the left renal vein branch.

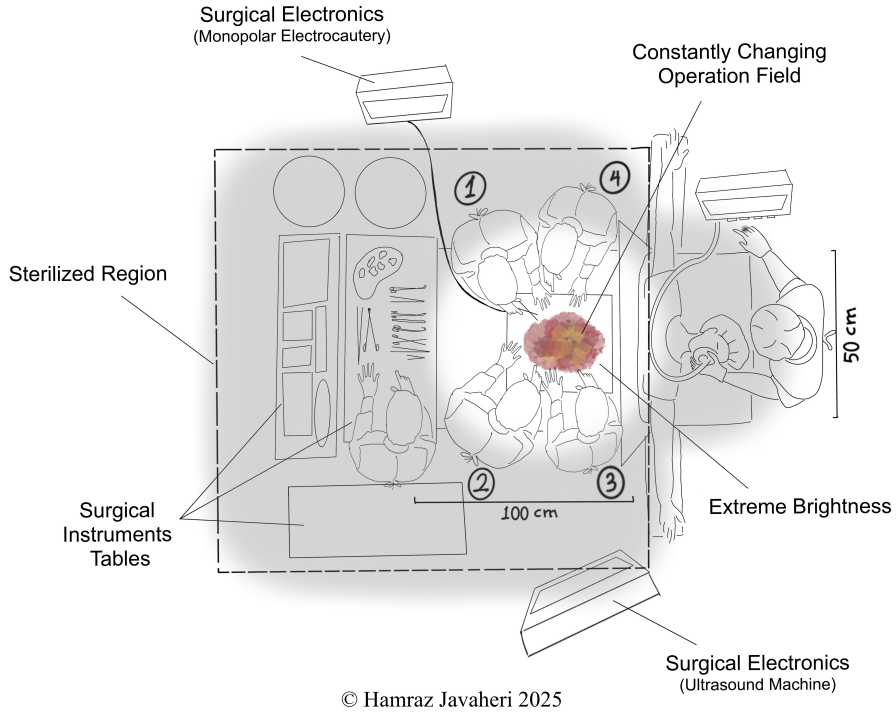


Figure 5.12.: Drawing depicting the environmental constraints within an operation room and the placement of the medical staff around the table. The surgeons are numbered based on their role in the surgery, with number 1 being the first lead surgeon, 2 the second leading surgeon, and 3,4 assistant surgeons.

Interaction Method We chose a fully voice-controlled approach for working with the in-situ virtual 3D model. This decision was driven by environmental and user constraints. The spatial limitations around the operating table made mid-air hand gestures impractical, as there was limited space for reaching or extending (Figure 5.12). Additionally, hand tracking near the surgery table faced challenges due to the patient's exposed abdomen and other surgeons' hands. Hence, to prevent accidental input, we disabled any interaction with the 3D model using hand gestures.

Furthermore, surgeons frequently wear loupe (magnifying) glasses during surgery for a magnified view of the operation site. The shape and bulkiness of these glasses, along with their positioning on the surgeons' faces, made it difficult to obtain reliable eye-tracking data. Given the stringent hygiene requirements, using external mice or keyboards was also impractical. Consequently, our system design relied entirely on voice control and speech commands.

Our considerations regarding the intuitiveness of the interaction method led us to choose specific voice control keywords that include the exact names of the reconstructed segments (Figure 5.9) identical to those used in ARAS preoperative planning for interaction with 3D visualization (Table 5.2). This approach allowed surgeons to interact with certain segments of the 3D model through-

out the surgery, such as enabling or disabling the segment-based visualization using only voice control without facing memorization issues. Additional to those keywords used in preoperative mode, we implemented 3 more control commands that enable interaction with marker tracking (Table 5.4).

Table 5.4.: Additional implemented voice keywords for intraoperative ARAS and their functionalities

Type	Voice Keywords	Functionality
Control commands	Freeze	Freezes the 3D model position and orientation and disabling the marker tracking
	Marker tracking	Enables marker tracking
	Reset	Resets the app including the 3D model position and orientation relative to sternum marker

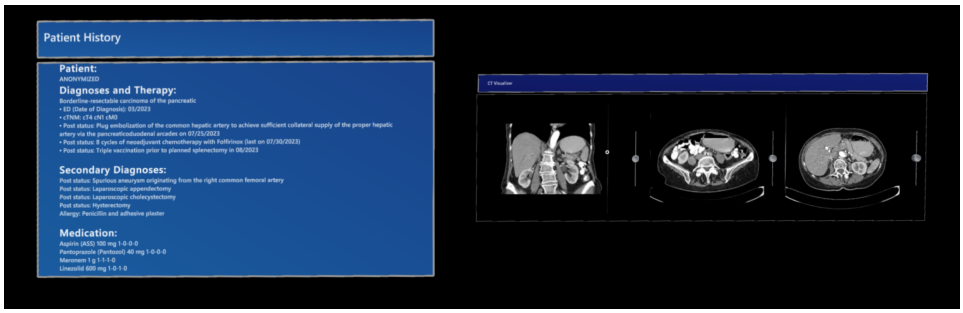
Supportive Data Visualization

A secondary feature of the ARAS intraoperative mode was to enable patient data visualization that could be controlled by surgeons without interfering with sterilization rules. Typically, this requires remote control by a third person in the operating room due to sterilization constraints, which reduces time efficiency when accessing the desired patient data or specific DICOM[7] slice. The same patient data as preoperative planning mode were included in the intraoperative mode. These data included the patient's diagnostic history and DICOM images, which are routinely used to plan the surgical procedure and can also play a significant role in decision-making during the operation. Therefore, both types of data were made available to the surgeons as visualizable features within the ARAS intraoperative mode. The patient's diagnostic history was presented as a scrollable text panel, and the three different phases of DICOM images captured before the surgical procedure were displayed in a scrollable DICOM viewer panel as detailed in Table 5.5.

Interaction Method Our investigation into finding the optimal interaction method for supportive data visualization revealed that it was less constrained compared to in-situ visualization, as virtual object placement was not as sensitive and could be positioned outside the operating field. We used the same approach as for the in-situ visualization to initialize data visualization with speech commands and employed a combination of mid-air hand gestures and speech commands to further interact with these objects. Due to equipment such as loupe glasses worn by the surgeons, we could not use eye-tracking data for gaze purposes; therefore, we opted for hand gaze instead of eye gaze.

[7] DICOM is a technical standard for the digital storage and transmission of medical images and related information.

Table 5.5.: The characteristics of in-situ visualization and the feature set provided by the ARAS software.



Supportive Data Visualization

Data type	DICOM images
	Patient history
Placement method	Automatic initial placement at closest ($> 1m$) non- colliding gazed position
	Manual repositioning
Features	Scrollable 3x DICOM visualizer
	Scrollable patient history panel
Interaction method	Mid-air hand gestures
	Keyword-based voice control

Surgeons could initiate patient history or DICOM images by using medical terminology associated with these data. The identical commands from ARAS preoperative planning were chosen (Table 5.2) for interacting with supportive data visualization. For instance, for the visualization of patient history, we used synonyms such as “diagnosis”, “medication”, and “patient history”. For the DICOM viewer, we used related medical terminologies such as “Tomography”, “CT”, “Computed Tomography”, and “CT scans”. These keywords were chosen based on input from the participating surgeons to ensure their relevance to their clinical routine (Table 5.2).

After the detection of relevant voice keywords, the interactable objects were placed at the closest ($> 1m$) non-colliding position in the direction of the user’s head gaze. If the placement of the objects was not suitable, surgeons could move or scale the virtual objects according to their preferences (see an example snapshot in Figure 5.13) using hand gaze and pinch gestures, speech commands, or a combination of both. For scrolling through DICOM slices and the patient history panel, we also employed speech commands to navigate and pause scrolling. Additionally, we included scroll buttons that could be used to scroll through the slices or the patient history panel using hand gaze and pinch gestures.

5.3.3. Companion Application

To automate and simplify the process of updating files for each patient by surgeons, we developed a local file transfer system using a Wi-Fi network. This system securely transmits anonymized patient data to the ARAS main software, as illustrated in Figure 5.14. The transmitted files include the patient’s

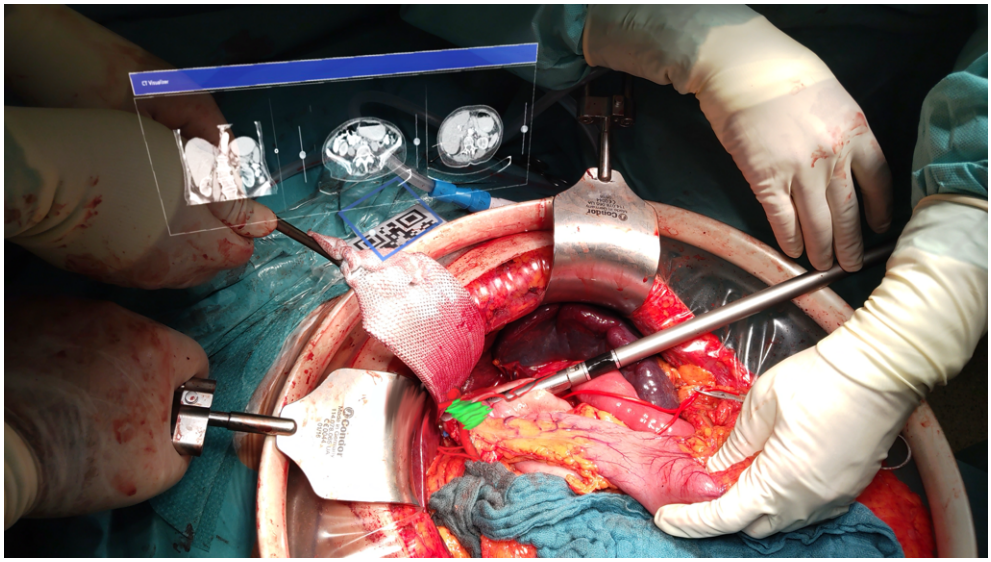


Figure 5.13.: A snapshot showing manual placement of scaled CT slices near the operation field done by the surgeon during the operation for a closer and continuous observation.

history, recent CT slices (captured in less than two weeks), and 3D models of the patient's vascular system and tumor, which are reconstructed from the CT scan slices using MeVis Liver Suites software [178].

The process of reconstructing patient-specific 3D objects is undertaken by surgeons, leveraging their expertise to identify structures from CT images. For each patient, the user (the surgeon) uses our custom desktop file transfer application to select patient files from a personal computer. Upon file selection confirmation, these files are encrypted, converted into a Unity asset bundle, and stored in a predefined directory. Additionally, the file transfer application establishes a local HTTP server over a Wi-Fi network. After the Unity asset bundle, containing the patient files, is stored, the surgeon can request an upload (for new files) or an update (for previously shared files) to make these files available for use and visualization in the ARAS software. On request, the server accesses the Unity asset bundle in the local directory and signals an update. The ARAS software, monitoring for this signal, automatically downloads and decrypts the updated patient files upon detection. This streamlined method facilitates real-time updates of patient files without the need to rebuild the software application for each case, thereby easing the integration of the application into clinical routines.

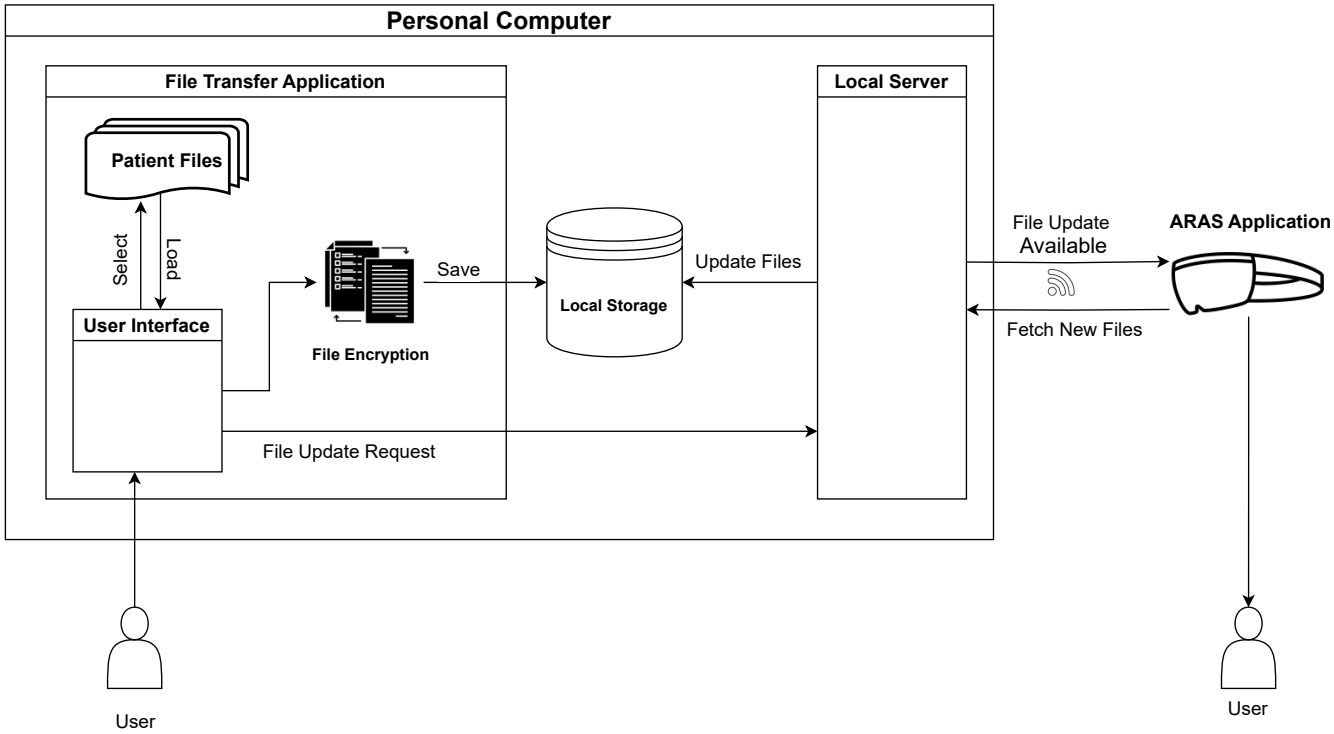


Figure 5.14.: The file transfer system used to share the patient files from a personal computer containing the files to the ARAS software on the HoloLens

5.4. Clinical Trials

Our system evaluation was conducted in two distinct phases. In the initial phase, the system was extensively tested by surgeons within a laboratory environment. Once it was confirmed that no further modifications were required, a clinical trial was initiated to validate our approach and system design in an ecologically valid setting. Clinical trials involved the preoperative use and intraoperative evaluation of ARAS in patients with underlying (borderline) resectable pancreatic tumors who required various types of pancreatic resection. It took place at Saarbrücken Klinikum hospital, a certified hepatopancreatobiliary center.

During the first five surgeries, the system underwent iterative refinement as real-world challenges emerged, issues that could not be fully simulated in the lab. One notable issue was the system's initial reliance on a single registration anchor point (the xiphoid process of the sternum). In occasional cases, surgeons required additional space in the patient's abdominal area, necessitating manipulation of the rib cage. This movement compromised the accuracy of the registration point and resulted in methodological failure.

In response, we introduced a secondary registration anchor (the vena cava) to ensure robustness and usability of the system even when the rib cage was manipulated. This iterative refinement phase was a critical step in resolving methodological limitations and ensuring the system's reliability and usability in real-world surgical scenarios.

Throughout the clinical trial, surgeons bore full responsibility for the safe use of the system during patient procedures, strictly adhering to ethical standards and ensuring patient safety at all times.

By the conclusion of the fifth surgery, the system had reached a stable state, and no further refinements or modifications were necessary.

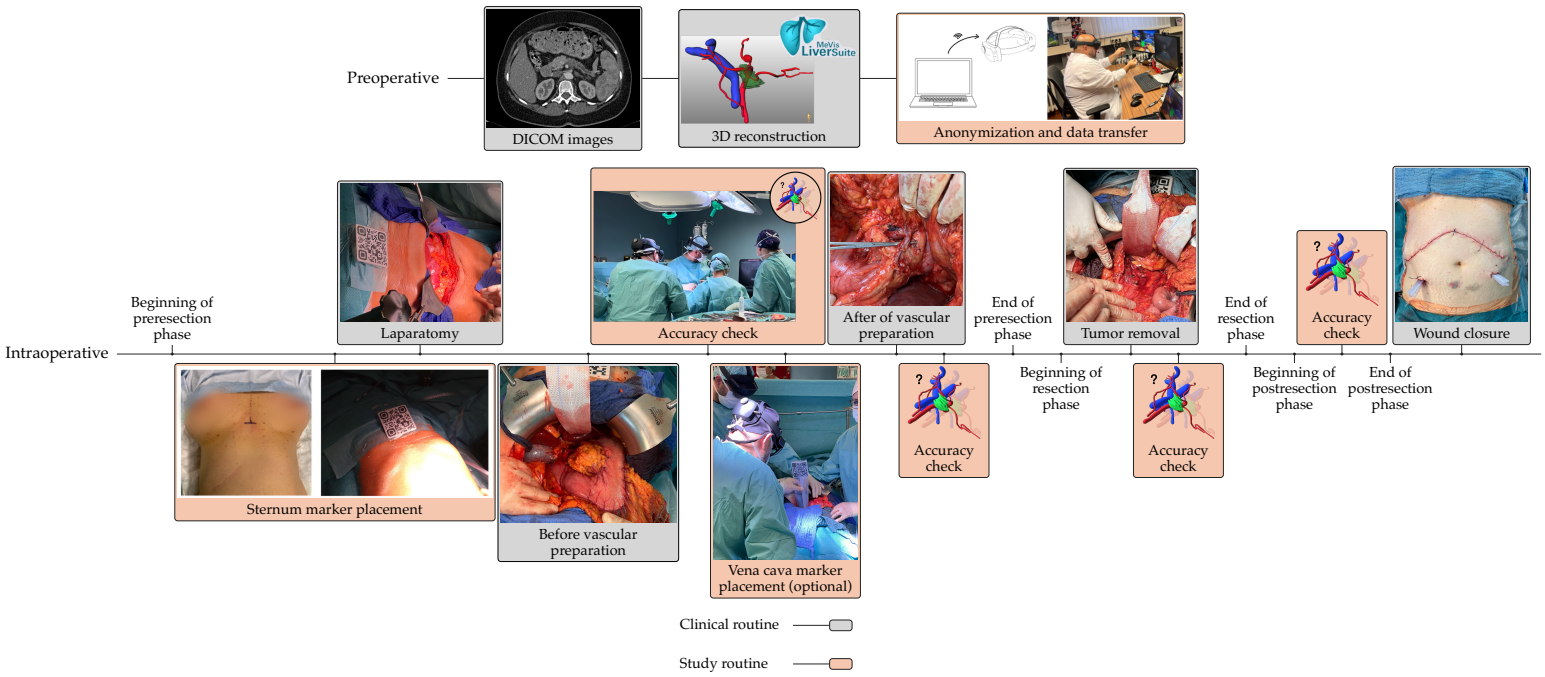


Figure 5.15.: The timeline showing the preoperative and intraoperative stages of the clinical trials, highlighting both routine clinical steps and specific study procedures required for ARAS.

5.4.1. Procedure

Patient inclusion in clinical trials began only after obtaining informed consent from each participant (Figure 5.1 clinical trials, patient inclusion phase). The process involved two steps: preoperative preparation and intraoperative evaluation of ARAS (Figure 5.15). During the preoperative phase, in addition to the standard clinical routine, a surgeon transferred anonymized patient data to the ARAS software after performing 3D reconstructions of the patient's vital structures from CT images (Figure 5.15, preoperative phase). Later, surgical planning was done using the ARAS planning mode.

Four surgeons participated in each surgery session. During surgery, at least two surgeons used the ARAS system continuously, while other surgeons used it briefly at each stage of the surgery, with the option to extend their use if desired. Due to the extreme expertise and specialization required to perform such surgical intervention, the surgeons were assigned based on the surgery plan and the hospital schedule.

As depicted in Figure 5.15, the intraoperative surgery session began with a surgeon identifying the anchor point for the marker on the patient's sternum and attaching the 3D marker. The surgeon located the xiphoid process of the sternum and the jugular notch by hand, as these anatomical landmarks are easily identifiable through the skin. A line was then drawn from the jugular notch to the xiphoid process using a surgical marker, and the registration point for the marker was marked at the xiphoid process. To confirm the accuracy of this point, the distance from the marker to the jugular notch was measured with a ruler and compared to the measurements from the CT images. The marker was then positioned flat, aligned at a 90-degree angle to the drawn line, and attached so that the end of the marker aligned with the end of the xiphoid process (Figure 5.11). Finally, a transparent surgical tape was applied to secure the marker in place without obscuring its features (Figure 5.15, intraoperative phase). This process took on average less than 3 minutes. Towards the end of preresection phase after mobilization of the pancreas head and vascular preparation, surgeons had the option to use the second marker - vena cava marker (Figure 5.11 (B)) - if the sternum was mobilized due to the surgical procedure.

At the beginning and end of each phase of the surgery (preresection, resection, and postresection) all four surgeons verified the accuracy of the tracking and the registration of structures which are not fully visible during various phases of surgery including arteries, veins, and the tumor (only in the first two phases before and during tumor removal) using an ultrasound probe (Figure 5.15 intraoperative accuracy check) and by palpating the arterial pulse [8] Ultrasound was selected due to its routine use in surgical procedures, thereby preventing any extra workload or significant delays in the operation. This choice also allowed surgeons to make informed decisions regarding reliance on

[8] Palpating the arterial pulse refers to the process of using one's fingers to feel the arterial pulse, which is the rhythmic expansion and contraction of an artery as blood is pumped through it by the heart. This is a common clinical practice used to assess the cardiovascular system.

AR visualization for guidance, while simultaneously evaluating the registration accuracy.

Surgeons could decide to discontinue using the system under the following circumstances: if its continued use compromised the progression of the surgery or the patient's safety, adversely affected the surgeons' health, comfort, or performance, if the method lost accuracy, if technical issues occurred, or if the system's assistance was no longer necessary.

Immediately after the surgical procedure, all four participating surgeons completed a custom questionnaire consisting of ten 5-point Likert scale questions (Table 5.6) assessing the accuracy of registration for each virtual structure type (arteries, veins, and tumor) during each phase of the surgery (preresection, resection, and postresection). Additionally, an observer researcher actively participated in all surgical sessions, recording the surgical procedure and ARAS usage durations, taking notes, and observing the surgeons' interactions with the system, as well as identifying any issues or challenges they encountered while using the system. After each surgical session, the same observer researcher conducted interviews with the surgeons who used ARAS continuously throughout the surgery.

Table 5.6.: 5-point Likert questionnaire on registration accuracy across surgery phases

	Question	Very Poor	Poor	Average	Good	Excellent
Preopertive Planning Phase						
1.	Usability of ARAS in surgery planning	□	□	□	□	□
2.	Applicability of preoperative plan with ARAS to intraoperative session	□	□	□	□	□
Intraoperative Preresection Phase						
3.	Registration accuracy of virtual tumor location	□	□	□	□	□
4.	Registration accuracy of virtual arterial location	□	□	□	□	□
5.	Registration accuracy of virtual venous location	□	□	□	□	□
Intraoperative Resection Phase						
6.	Registration accuracy of virtual tumor location	□	□	□	□	□
7.	Registration accuracy of virtual arterial location	□	□	□	□	□
8.	Registration accuracy of virtual venous location	□	□	□	□	□
Intraoperative Postresection Phase						
9.	Registration accuracy of virtual arterial location	□	□	□	□	□
10.	Registration accuracy of virtual venous location	□	□	□	□	□

5.4.2. Accuracy Check

All surgeons who participated in the system's evaluation during clinical trials were fully informed and aware of the experimental nature of the method and the device used. Subjective measures (RQ2 method) were chosen due to the lack of a reliable objective method for calculating registration errors between non-visible internal structures and their visualization in 3D coordinates on a head-mounted display during surgical procedures.

While objective methods, such as fiducial registration error (FRE), which measures the registration error by calculating the distance between a fiducial and its corresponding point after registration [172], are well-established in image-guided surgery [164] due to their simplicity and speed in estimating target registration error (TRE) for non-visible structures [56], their application in head-mounted displays for surgical contexts remains underexplored. Furthermore, previous studies have shown that TRE is not always correlated with FRE [56, 83], and FRE is considered an unreliable indicator of registration accuracy [84]. Therefore, in the context of measuring 3D registration error for specific non-fiducial segments visualized in an AR head-mounted display, FRE is an unreliable predictor of TRE. Although landmarks not visible to the naked eye can be identified using imaging devices such as ultrasound, measuring the 3D registration error between a head-mounted display and medical images for such landmarks is prone to error. Thus, any metric accuracy report based on 3D calculations from landmarks identified in medical images would also be unreliable. Given that the participating surgeons (Table 5.1) were experts in their field (average surgical experience = 16 y, SD = 8.14 y), their subjective evaluations are considered more reliable and informative in this context due to their expertise and understanding of clinical requirements.

5.4.3. Patients and Types of Operations

The surgeons (Table 5.1) evaluated the ARAS system for open PS across three different types of operations, involving 20 patients (Table 5.7): the pancreaticoduodenectomy [9], total pancreatectomy[10], and distal pancreatectomy[11] In both the pancreaticoduodenectomy procedure and total pancreatectomy, full mobilization of the pancreatic head is required, whereas in distal pancreatectomy, organ mobilization is minimal during the surgical procedure. One of the significant challenges of using AR and preoperative data during surgery

[9] The Whipple procedure, also called pancreaticoduodenectomy, is an operation to remove the head of the pancreas.

[10] A total pancreatectomy is a surgical procedure where the entire pancreas is removed, often along with other nearby organs or tissues. These can include the duodenum (part of the small intestine), the gallbladder, parts of the bile duct, the spleen, and sometimes part of the stomach.

[11] A distal pancreatectomy is a surgical procedure where the tail and sometimes part of the body of the pancreas are removed, while the head of the pancreas is left intact. In many cases, the spleen is also removed during the procedure because of its close proximity to the pancreas. Unlike the pancreaticoduodenectomy procedure, a distal pancreatectomy focuses on the left side (distal portion) of the pancreas.

Table 5.7.: Demographics and clinical data of patient cases during the clinical trials. We report the biological sex of the patients.

Age	Sex	Performed surgical procedure
48	Female	Distal pancreatectomy
82	Female	Distal pancreatectomy
73	Female	Distal pancreatectomy
59	Male	Pancreaticoduodenectomy
69	Female	Pancreaticoduodenectomy
80	Female	Pancreaticoduodenectomy
68	Male	Pancreaticoduodenectomy
76	Female	Pancreaticoduodenectomy
65	Female	Pancreaticoduodenectomy
62	Male	Pancreaticoduodenectomy
76	Male	Total pancreatectomy
62	Male	Distal pancreatectomy
73	Female	Pancreaticoduodenectomy
59	Male	Total pancreatectomy
77	Male	Pancreaticoduodenectomy
68	Female	Pancreaticoduodenectomy
70	Male	Pancreaticoduodenectomy
55	Male	Pancreaticoduodenectomy
66	Female	Pancreaticoduodenectomy
48	Female	Pancreaticoduodenectomy

is maintaining registration accuracy despite the intraoperative mobilization of organs. By including these varied cases, we were able to comprehensively evaluate the system under different conditions.

All patients were fully informed about the study, and were included in the study only after a thorough explanation of the study and the collection of informed consent. Patients were informed about the potential risks, benefits, and alternatives to participating in the study, ensuring that their participation was voluntary and based on a clear understanding of the procedure.

5.5. Results

This section presents the outcomes of system evaluation during the clinical trials and the overall study results. It begins by detailing the quantitative findings (Section 5.5.2 and Section 5.5.3) and qualitative insights (Section 5.5.4 and Section 5.5.5) from the clinical trials.

5.5.1. ARAS Planning

The ARAS planning mode was highly appreciated by all surgeons, receiving an average score of 5 (SD = 0) on both scales in the questionnaire. The ability to visualize and manipulate the patient's 3D model and perform planning in 3D rather than 2D was considered highly beneficial in all cases.

Interestingly, in two cases, ARAS planning led to a different surgical plan than what would have been decided based solely on CT images. Patient 2 and Patient 3, 82- and 73-year-old female patients, respectively, were diagnosed with cystic tumors in the body of the pancreas, both exhibiting features suggestive of malignancy. As a result, surgical resection of the lesions was deemed necessary. Preoperative planning was performed using CT images and a reconstructed 3D model of ARAS. Upon reviewing the CT images, spleen-preserving distal pancreatectomy was initially planned for patient 2. However, the reconstructed 3D model revealed a highly tortuous course of the splenic artery along with the proximity of the tumor to the splenic vein. Consequently, the ARAS assessment indicated that spleen preservation would be challenging. For patient 3, ARAS-guided planning led to the decision to perform an extended distal pancreatosplenectomy.

5.5.2. ARAS Intraoperative: Surgical Procedure and ARAS Usage

The surgeries lasted an average of 245.9 (SD = 49.23) minutes. Out of this, 38.5% of the time was spent on the preresection (vascular preparation) phase, 9.6% on the resection phase, and 51.90% on the postresection phase. The longest continuous duration of ARAS usage across all sessions was on average 122.9 (SD = 35.65) minutes for an average 50.37% (SD = 11.31) of the total surgery time. No complications associated with ARAS occurred during and after the operations. In all sessions, surgeons reported discontinuing the use of the ARAS for four main reasons: (1) the vital structures were already exposed and no longer needed AR support (77.08%); (2) discomfort due to the device's high temperature, sweat, and neck pain (16.67%); reduced model accuracy due to surgical manipulations (4.17%); and battery-related issues (2.08%). Notably, none of the surgeons reported discontinuing the use of the ARAS due to patient safety concerns.

5.5.3. ARAS Intraoperative: Registration Accuracy and Usability

The questionnaire results on intraoperative registration accuracy (Table 5.6) of the virtual models (incl. arteries, veins, and tumor) yielded overall average scores of 4.6 (SD = 0.56), 4.0 (SD = 1.02), and 3.60 (SD = 0.78) on a 5-point Likert scale for preresection, resection, and postresection phases, respectively. Figure 5.16 depicts these results categorized by operation phase and structure.

The UMUX-Lite score of ARAS using a 7-point Likert scale was 6.50 (SD = 0.35), and the corresponding SUS score, predicted using a regression equation based on the two UMUX-Lite items [156], was 82.48. According to the rating scale by Bangor et al. [18], this places ARAS in the "good" (close to "excellent")

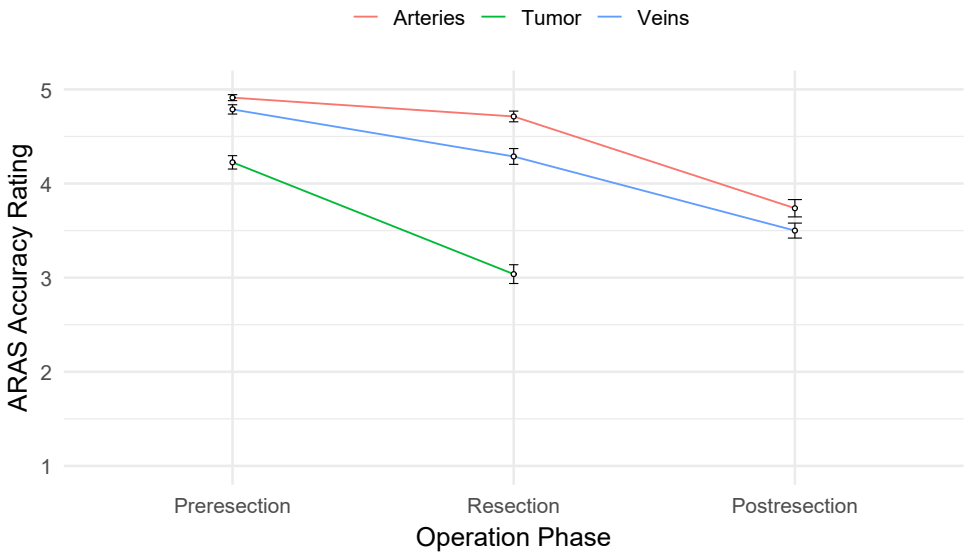


Figure 5.16.: Registration accuracy of patient-specific virtual models during each phase of the operation using a 5-point Likert scale. Error bars indicate standard error.

category for system usability.

5.5.4. Postoperative Interviews

Through our thematic analysis of the postoperative interviews (Section 5.2.5), we developed five main themes:

Clinical Accuracy and Usability A major point frequently emphasized by surgeons was ARAS's accuracy in providing visualizations and guidance, particularly in overlaying crucial structures and offering visual assistance during surgery. Surgeons notably valued the system's capacity to help them precisely locate and identify critical structures during intricate dissections. This sentiment was expressed by a second lead surgeon as follows:

> *"There was a moment when we didn't know where the most important artery is. The chief surgeon pointed to the location where the artery was visualized in HoloLens. And then we dissected along that point, and we saw that the artery was exactly there! It was truly amazing!"* (S5)

Additionally, participants reflected on the usability of the voice user interface for interacting with the 3D model, reporting occasional performance degradation due to the system's need for precise keyword articulation, warranting more robust voice control methods less affected by environmental noise. One assistant surgeon encapsulated this issue with the following sentiment:

> *"speech commands worked fine, but the issue with the speech commands is sometimes when people are talking around the table, you have to say the word a couple of times until the system detects it."* (S6)

Aid in Decision-Making Another emphasized topic during the interviews was how the in-situ navigation provided by ARAS significantly aided surgeons in decision-making during critical moments of the operation. Two lead surgeons highlighted this:

> *"When we were dissecting along the SMA (superior mesenteric artery), we were not sure where it really was, or we couldn't really find it. So that was extremely helpful to really be sure that it's there."* (S4)

> *"It was one very critical situation in the operation. We could have ruined the whole case. But I had a view before in ARAS, and I had an idea where the SMA was. It was quite dangerous at that moment. But we saved everything. So I like it a lot. I think it's a great help, especially in these situations."* (S1)

Educational Efficacy and Training Enhancement Moreover, surgeons consistently emphasized that ARAS not only aided in decision-making during surgeries but also proved to be an invaluable tool for enhancing the education and training of junior surgeons, especially for complex procedures. The assistant surgeon, who had less experience than the lead surgeons, reflected on this:

> *"When I participated in this type of operation for the first couple of times, I never understood what the chief surgeon was doing, because I didn't know where he was going in, I didn't know which vessels he wanted to focus on. You also see that sometimes even senior surgeons misrecognize the vessels during the operation. Even when you are experienced, it's very difficult to understand the anatomy of the patient during surgery. With this system (ARAS), I think I understood this operation better and could clearly see where the vessels are."* (S6)

Technical Reliability and Possible Improvements The combined use of the xiphoid process of the sternum and an occasional anchor point on the vena cava was generally regarded as a reliable method, especially during the initial and most critical stages of surgery. However, in more advanced stages, such as after tumor removal or during the pancreaticoduodenectomy procedure after mobilizing the pancreatic head (resection phase), the accuracy of some 3D models was compromised, as it was also reflected in accuracy measurements (Figure 5.16). This indicates a potential need for real-time data integration to

maintain model accuracy throughout the surgery. However, it was also noted that at these later stages, surgeons typically require less navigational assistance since the critical structures are already exposed and visible, as highlighted by a surgeon.

> *"After we mobilized everything in the head of the pancreas, the virtual tumor position was not correct anymore. (...) But that was not important for us at that stage, as we could already see the tumor, and it would be removed with half of the pancreas anyway."* (S3)

Ergonomics Finally, surgeons addressed the system's ergonomics, noting that they experienced no issues with the comfort or overall usability of the HoloLens 2 during surgery. No surgeons reported any side effects associated with head-mounted wearables, such as motion sickness, pain, or nausea.However, for some of those who also used magnifying glasses occasionally faced discomfort and problems with the AR view when combining the magnifying glasses with the HoloLens. A lead surgeon commented on this:

> *"The device is okay. There's no problem with wearing it. You can wear it for a while without disturbance. But for me, the one problem is the loupe glass. It's quite in the way and causes quite a difference in AR view."* (S1)

Furthermore, some surgeons reported occasional discomfort, such as extensive sweating and neck pain, while wearing the HoloLens 2 for an extended time. An assistant surgeon addressed this issue, stating:

> *"The device got too hot during the operation, I just couldn't wear it anymore as I started to sweat a lot."* (S7)

5.5.5. Intraoperative Observations

After thematic analysis of the observational data (Section 5.2.5), we developed three main themes.

Interaction Challenges Throughout the clinical trials, we noticed a gradual reduction in the difficulties associated with using the device and system, indicating an adaptation period and a learning curve. This is likely due to the unfamiliarity of the technology and interaction methods, such as mid-air hand gestures and voice control, particularly in the medical field.

Given the high stress levels and cognitive demands on surgeons during surgery, any unnecessary interaction or additional functionality led to confusion, especially as the number of features increased. This was particularly evident when more detailed voice control options were introduced to manage marker tracking modes and data visualization panels.

These observations underscore the need to further adapt interaction methods for use in high-stress, time-sensitive environments. Additionally, some surgeons occasionally struggled with keyword-based voice control, requiring multiple attempts due to incorrect word recognition, often caused by operating room noise and user accents. Therefore, methods for the integration of a more natural communication system that allows for intuitive interaction beyond keyword-based voice control should be explored.

Emergence of User Roles The presence of multiple ARAS users during surgery led to the development of distinct roles among them. It was noted that the assistant who continuously used ARAS often took on the role of navigator, actively monitoring the operation through the AR glasses and alerting other surgeons about hidden or hard-to-see structures. Meanwhile, the lead surgeons focused on the real environment of the surgery and referred to the virtual overlay as needed. This observation not only underscores the natural evolution of user roles but also suggests a key consideration for future system design and development, where different modes could be tailored based on users' roles and profiles in the surgical process.

User Acceptance Throughout the trials, a notable increase in user trust and acceptance of ARAS was observed, particularly among surgeons who initially interacted with the system only briefly for accuracy checks. At the outset, these surgeons preferred not to use ARAS for extended periods, favoring an unobstructed view and full attention on the surgical procedure. However, as the trials progressed and the system's usability became evident, they demonstrated a greater willingness to use ARAS for longer durations. This shift in behavior highlights a growing trust and acceptance of the technology over time. The users' initial skepticism can be overcome as they become more familiar with and reassured by the system's performance. This underscores the importance of demonstrating the practical benefits and reliability of new technologies through comprehensive training, clear communication about the system's capabilities and limitations, and ongoing support to gain user acceptance.

5.6. Discussion

Our discussion provides a holistic view of the presented work, while addressing specific results from the clinical evaluation, highlighting challenges encountered, and outlining implications for future research.

5.6.1. Design Implications for Interaction with AR-based Systems in Open Surgery

Our study highlights the crucial role of a "less is more" approach in time-critical surgical settings, particularly in enhancing the perceived ease of use

and effectiveness of AR navigation systems. The findings underscore the importance of delivering only the most relevant and essential information to surgeons at the moment it is needed. This aligns with established HCI principles, such as the concept of details on demand ([235]). We recommend that future designers **prioritize minimalistic data visualization that presents only essential information by default, while allowing access to more detailed data as needed.** This approach helps prevent cognitive overload and promotes clarity in high-pressure clinical environments.

Consistent with previous research on various interaction modalities [106, 124, 176], our investigation revealed that the benefits of **touchless interaction methods for the surgical domain are circumstantial rather than absolute**. When designing for such a critical field, developers must carefully consider domain-specific constraints. for each system feature. For example, the interaction method should be determined based on factors like the proximity of the object, the sensitivity of the data, and the user's capabilities in specific situations. In this context, introducing system restrictions such as separating interaction methods for near and far objects, or restricting the use of input modalities like voice, hand, or gaze, could help prevent unintended interactions, especially when displaying highly sensitive data such as navigational 3D overlays.

Data sources or input modalities that perform unreliably due to interference with the user's equipment should be avoided. In our study, we found that relying on speech commands and hand tracking through the HoloLens cameras and microphones, combined with specific feature-based restrictions, provided an optimal interaction approach. Unlike in laparoscopic surgeries, where eye gaze could be employed for various tasks [212], our preliminary experiments identified limitations with the HoloLens 2's embedded eye tracker. The interference from bulky loupe glasses, commonly used in open surgeries to enhance the visualization of delicate structures, reduced eye tracking reliability. As a result, we chose to exclude eye tracking data to avoid unintentional interactions. Future research could explore seamless integration of magnified views within see-through AR devices, enabling reliable use in open surgery to mimic the functionality of loupe glasses.

Lastly, our observation revealed that multiple ARAS users naturally developed distinct roles during surgery. Assistant surgeons using ARAS often acted as navigators, monitoring the operation through the AR display and alerting lead surgeons to hidden structures, while the lead surgeons focused on the surgery itself. This suggests that **future surgical AR systems should consider role-based functionality**, with tailored modes for different users, such as a "navigator" mode for assistants and an "overview" mode for lead surgeons, to enhance system efficiency and usability. Further research on such features that shape engagement and team collaboration while using AR devices in the surgical environment should be conducted for a better understanding of the system design's impact on communication dynamics, decision-making efficiency, and overall surgical outcomes.

5.6.2. Feasibility and Accuracy of ARAS Intraoperative

ARAS is a feasible solution for continuous and accurate in-situ visualization of patient anatomy during open PS without interfering with the procedure. **The questionnaire results and interviews confirmed the high registration accuracy achieved by our dual-layer marker-based registration approach during the phases with the greatest need for ARAS: the preresection and resection phases** (Figure 5.7).

In line with work by Onda et al. [199], our study revealed that the 3D models were accurately registered during the initial phases of the surgery but declined as the procedure progressed. This effect is particularly noticeable in tumor registration due to relational deformations occurring during surgery as a result of surgical manipulations. Unlike vessels, which are typically anchored to the abdominal cavity behind the pancreas, hence remaining stable, tumors are often attached directly to the pancreas and move along with it during the preresection phase. While the registration accuracy remains high for vessels, the registration of the tumor declines as more manipulations are performed during the surgery.

These relational deformations are primarily caused by surgical manipulations and disrupt the spatial relationships between structures compared to their preoperative state. This explains the decline in accuracy, particularly after the resection phase (Figure 5.16). As such, our approach of leveraging **preoperative data for intraoperative surgical navigation is limited by surgical shifts and structural alterations in the target structures**. Its accuracy relies on the constant spatial relationship between all 3D segments throughout the surgery. Unfortunately, this limitation cannot be addressed unless each target segment is tracked individually, which is not feasible, as most vital structures remain hidden until the resection phase. While deformation estimation approaches [94] might reduce registration errors caused by relational deformations, they are still prone to errors caused by surgical alterations, as they rely on deforming the preoperative data.

Likewise, **continuous intraoperative usage of real-time data sources such as ultrasound, MRI, or CT to update the data remains challenging for open surgery.** These data acquisitions are often manual and time-consuming. Unlike methods used in laparoscopic surgeries [226], benefiting from real-time data from an ultrasound probe attached to the laparoscopic camera, the practicality of such an approach for open surgery is rather limited due to the manual manipulation required to hold the ultrasound probe throughout the surgery. Likewise, other real-time data sources, such as MRI or CT, delay the surgical procedure and cause unnecessary radiation exposure. In critical fields like tumor removal surgeries, any delay in the surgical procedure may cause complications and worsen patient outcomes [43], limiting the usage of medical imaging as a real-time data source.

In pursuit of avoiding delays in the surgical procedure, we also defined the total surgery time as one of our evaluation metrics, in addition to registration accuracy. **The average surgery time with ARAS remained within previously reported surgical ranges** for the same type of procedure [54, 207,

284], demonstrating its feasibility for intraoperative use without introducing any delays to the surgical process.

Consistent with previous studies demonstrating the effectiveness of in-situ AR navigation for PS using display-based AR [168, 195, 252], our research also validated and further improved this approach by utilizing a wearable head-mounted AR device in clinical trials with a larger sample size. **Our hands-free solution overcame the limitations of hand-held devices by enabling continuous use throughout the surgery.**

Throughout 20 surgeries, we validated our system usability and accuracy across three different surgical procedures. In line with the findings of Marzano et al. [168], **we proved the feasibility of our approach for surgical procedures characterized by increased intraoperative mobilization** due to surgical manipulations, such as pancreaticoduodenectomy, and also for other types of PSs (Table 5.7). Our study results showed no significant difference in the reported registration accuracy for different types of operations, demonstrating ARAS's usability across different PSs.

5.6.3. Ethical Considerations

Ethical adherence in this field is critical for maintaining trust in medical research and innovation. By following ethical guidelines, researchers and practitioners demonstrate their commitment to conducting research that prioritizes patient safety and well-being over the pursuit of technological advancement.

In this study, the adherence to ethical guidelines was our primary concern throughout all phases of the study. We ensured that the introduction of the AR system did not compromise the safety of patients or surgeons, nor the integrity of the surgical process. We made sure of full human decision-making in surgical procedures, and that the AR system never replaced the surgeon's judgment. Surgeons maintained full control over the system, ensuring complete adherence to the Fundamental Principles of Ethics [260].

Our key takeaways from this study regarding ethical considerations for future studies are as follows: **The system should only serve as a supplementary tool to assist the surgeon, augmenting rather than replacing their expertise**, by providing enhanced visualization and information to support decision-making. The surgeon must retain ultimate authority in decision-making, ensuring patient safety and adapting to the nuances of each individual case. Human intuition and experience should be valued as the final safeguard.

5.7. Conclusion

In this chapter, we detailed the successful design and implementation of ARAS, an AR assistance system for PS. Our research expanded the understanding of AR technology's application in the surgical field, moving beyond lab settings through clinical trials with patients. Accuracy measurements and postoperative surgeon interviews confirmed the successful implementation of our system,

especially during the crucial preparation and resection phases, where AR navigation assistance was most

Chapter 6

From Cognitive Relief to Clinical Benefit: Downstream Impact of Assistive Tools on Surgical and Patient Outcomes

Contents

As surgical procedures grow increasingly complex, the cognitive demands placed on surgeons, ranging from spatial reasoning and attention management to rapid decision-making, can directly influence clinical performance and patient outcomes. Assistive technologies such as AR and navigation systems are often introduced with the goal of easing these cognitive and motor burdens. While most studies focus on their immediate benefits, such as improved visualization or workflow efficiency, less is understood about how these upstream supports translate into downstream clinical value.

This chapter investigates the extended impact of assistive tools beyond the point of use. It explores how cognitive relief during surgery may lead to im-

proved surgical precision, reduced intraoperative errors, and ultimately better patient outcomes. By examining this link between human performance augmentation and clinical effectiveness, the chapter aims to contribute to a more holistic understanding of what makes assistive technologies truly valuable in high-stakes medical environments.

The content of this chapter covers the following journal publication: .

1. Javaheri, Hamraz, Omid Ghamarnejad, Rizky Widyaningsih, Ragnar Bade, Paul Lukowicz, Jakob Karolus, and Gregor Alexander Stavrou. "Enhancing Perioperative Outcomes of Pancreatic Surgery with Wearable Augmented Reality Assistance System: A Matched-Pair Analysis." Annals of Surgery Open 5, no. 4 (2024): e516.

This work was awarded the "Best Scientific Work" award by the Saarland Committee of Surgeons Conference in 2024, where it was first presented.

6.1. Background

This section provides background information and an overview of key medical terminology and guidelines relevant to this study and referenced throughout the chapter.

6.1.1. Overview of Pancreatic Surgery Types

Pancreatic surgery is selected based on the tumor's location, size, and relation to adjacent vessels and organs. The procedures vary in complexity and extent, ranging from local tumor excision to total removal of the pancreas. The table below summarizes the most common types of pancreatic surgeries used in clinical practice.

6.1.2. Terminology and Classifications

This section introduces essential clinical terms and classification systems used throughout this chapter. These concepts are key to understanding how surgical outcomes are assessed in pancreatic operations, particularly when evaluating the impact of ARAS on surgical performance.

Resection Margin Status (R-status)

In surgical oncology, the term "resection margin" refers to the edge of the tissue that has been removed. Its status indicates whether the tumor was completely removed.

- **R0**: No cancer cells are found at the resection margin. This indicates a complete (curative) resection.

Table 6.1.: Common Types of Pancreatic Surgery

Surgery Type	Main Indication	Part of Pancreas Removed
Whipple (Pancreaticoduodenectomy)	Tumors in the head of the pancreas	Head
PPPD (Pylorus-preserving Pancreaticoduodenectomy)	Head tumors, preserving stomach function	Head (pylorus preserved)
Distal Pancreatectomy	Tumors in the body or tail	Body and/or tail (often with spleen)
Total Pancreatectomy	Diffuse disease or multicentric tumors	Entire pancreas (plus surrounding organs)
Central Pancreatectomy	Small, benign tumors in the neck/body	Central part (neck/body)
Enucleation	Small, benign or neuroendocrine tumors	Tumor only, surrounding tissue preserved
Appleby Procedure	Locally advanced tumors with arterial involvement	Tail + celiac axis (arterial resection)
DPPHR (Duodenum-preserving Pancreatic Head Resection)	Benign lesions in the head	Head (preserving duodenum and bile duct)

- **R1**: Microscopic cancer cells are present at the margin, suggesting incomplete removal.
- **R2**: Visible (macroscopic) tumor remains. This is generally considered a palliative, not curative, procedure.

In this study, we distinguish between R0 and R1 resections. Achieving an R0 resection is a primary goal in pancreatic surgery because it is associated with improved long-term survival.

ISGPS Definitions for Postoperative Complications

The *International Study Group of Pancreatic Surgery (ISGPS)* provides widely accepted definitions for three common and critical complications following pancreatic surgery:

1. Postoperative Pancreatic Fistula (POPF) A pancreatic fistula occurs when pancreatic fluid leaks from the surgical site. The ISGPS categorizes this condition as:

- **Biochemical leak**: Elevated enzyme levels in drainage fluid, without clinical consequences.
- **Grade B**: Requires medical treatment or interventional drainage.
- **Grade C**: Severe, often requiring reoperation or intensive care, with significant risk of mortality.

2. Delayed Gastric Emptying (DGE) This refers to the stomach's inability to empty its contents in a timely manner after surgery.

- **Grade A**: Minimal symptoms, usually resolves without intervention.
- **Grade B**: Requires temporary reinsertion of a feeding tube or dietary restrictions.
- **Grade C**: Severe delay, often requiring prolonged hospitalization or additional treatment.

3. Postpancreatectomy Hemorrhage (PPH) This is bleeding that occurs after pancreatic surgery.

- **Grade A**: Mild bleeding with no clinical impact.
- **Grade B**: Requires blood transfusion or radiologic intervention.
- **Grade C**: Life-threatening bleeding requiring emergency surgery or intensive care.

Clavien–Dindo Classification of Surgical Complications

This classification system is used to grade the severity of all types of postoperative complications:

- **Grade I**: Minor deviation from the normal course, no intervention required.
- **Grade II**: Requires pharmacological treatment (e.g., antibiotics, transfusions).
- **Grade IIIa**: Requires surgical or radiological intervention without general anesthesia.
- **Grade IIIb**: Requires intervention under general anesthesia.
- **Grade IV**: Life-threatening complications requiring ICU care.
- **Grade V**: Death.

In this study, we used the following classification:

- **Minor complications**: Grades I to IIIa.
- **Major complications**: Grade IIIb and above.

Borderline Resectable Pancreatic Cancer

This term refers to tumors that involve nearby blood vessels (such as the superior mesenteric vein or artery) but may still be operable with advanced surgical techniques. These cases often require **neoadjuvant chemotherapy** (chemotherapy given before surgery) to improve the chance of a complete resection.

Appleby Procedure

A radical surgical technique is used when the tumor invades major arteries. It involves removing the distal pancreas along with the **celiac axis** (a major arterial branch) and is typically reserved for patients with advanced, but potentially resectable, pancreatic body tumors. This procedure requires detailed vascular planning and high surgical precision, making it a prime use case for ARAS.

6.2. Methodology

To assess the clinical and surgical outcomes of our AR Assistance System (ARAS), we conducted a comparative study following the completion of its clinical trial. This study aimed to evaluate the effectiveness of ARAS by comparing the surgical outcomes of patients who underwent pancreatic surgery with ARAS to those who underwent surgery without it.

The clinical trial was a single-center, prospective study initiated in September 2023. It enrolled 20 consecutive adult patients (aged over 18) diagnosed with (borderline) resectable pancreatic tumors and scheduled for elective open pancreatic resection. Upon the trial's completion, these prospective cases were compared to a retrospective cohort.

The retrospective data were sourced from a prospectively maintained database and included 180 consecutive patients who underwent pancreatic resection between January 2018 and September 2024. Exclusion criteria included patients who had undergone pancreatic enucleation, duodenum-preserving pancreatic head resection, urgent surgery, previous open upper gastrointestinal procedures, or laparoscopic pancreatic resections. After applying these criteria, 105 patients remained eligible for analysis.

To minimize selection bias and ensure comparability between groups, we performed a matched-pair analysis. The 20 patients who underwent ARAS-assisted surgery were matched in a 1:3 ratio with 60 patients who received standard surgical care. Matching variables were selected based on factors known to influence intraoperative outcomes, such as operative time and intraoperative blood loss. These variables included sex, age (± 5 years), body mass index (BMI, ± 5 kg/m^2), American Society of Anesthesiologists (ASA) classification, tumor type, and type of surgery (standard vs. extended pancreatic resection).

All procedures were conducted by the same experienced surgical team, and the study adhered to the principles outlined in the latest revision of the Declaration of Helsinki [13].

6.2.1. Measures and Collected Data

The primary aim of this study was to examine the effect of ARAS on intraoperative outcomes, including operation time (from incision to wound closure), excessive intraoperative bleeding, and red blood cell (RBC) transfusion. Ex-

cessive intraoperative bleeding was defined as blood loss exceeding 1,000 ml during the surgery.

The collected data included preoperative measures and also intra- and postoperative data. The preoperative data included patient demographics and clinical characteristics, were recorded, including age, sex, BMI, ASA classification, surgical indication, and any neoadjuvant chemotherapy received. Data collected during surgery included the type of procedure performed, any extended organ resections (such as colon, liver, stomach, adrenal gland, kidney, or small bowel), vascular resections (arterial or venous), and resection margin status (R-status). Following surgery, the length of stays in both the intensive care unit (ICU) and the hospital were documented. Postoperative pancreatic fistula 18, delayed gastric emptying 19, and postpancreatectomy hemorrhage (PPH) 20 were diagnosed and graded according to the criteria established by the International Study Group of Pancreatic Surgery (ISGPS). Postoperative complications were evaluated and categorized using the Clavien-Dindo classification 21. Minor complications included Grade I, II, and IIIa, while major complications were those classified as Grade IIIb or higher. Mortality was defined as any death occurring within 30 days postoperatively.

6.2.2. Patients and Demographics

A total of 80 patients were included in this study. Baseline demographic and clinical characteristics are summarized in Table 6.2. The average age of the patients was 67.4 ± 9.8 years, with 55.0% being female. Pancreatic adenocarcinoma was the predominant indication for pancreatic resection, accounting for 75.0% ($n = 60$) of cases. There were no significant differences in demographic variables, including age, sex, BMI, ASA class, and surgical indication, between patients who underwent pancreatic resection with or without ARAS. Notably, 20.0% ($n = 4$) of patients in the ARAS group were diagnosed with borderline resectable pancreatic cancer and received neoadjuvant chemotherapy, compared to 6.7% ($n = 4$) in the control group ($p = 0.085$). Two of these ARAS patients were diagnosed with pancreatic body cancer that had infiltrated the celiac axis and common hepatic artery. After receiving neoadjuvant therapy, they underwent an extended distal pancreatosplenectomy combined with en bloc celiac axis resection (Appleby procedure, Table 2).

6.2.3. Statistical Analysis

Statistical analysis was conducted using IBM SPSS Statistics for Windows, Version 29.0 (IBM Corp., Released 2022, Armonk, NY). Continuous data were expressed as means $\pm$ standard deviations, while categorical data were reported as frequencies and proportions. The comparison of continuous variables was performed using Student's t-test, and categorical variables were analyzed using the chi-square test or Fisher's exact test. A two-sided p-value of < 0.05 was considered statistically significant for all analyses.

Table 6.2.: Demographic and Preoperative Clinical Data

Variables	Total (n = 80)	ARAS (n = 20)	Control (n = 60)	*p*
Age (years)	67.4 ± 9.8	66.8 ± 9.7	67.7 ± 9.9	0.779
Sex				0.999
Female/male	44/36	11/9	33/27	
BMI (kg/m2)	25.4 ± 3.1	25.7 ± 2.9	25.4 ± 3.5	0.622
ASA class				0.999
<Class 3	12 (15.0%)	3 (15.0%)	9 (15.0%)	
≥ Class 3	68 (85.0%)	17 (85.0%)	51 (85.0%)	
Indication for surgery				0.999
Benign	8 (10.0%)	2 (10.0%)	6 (10.0%)	
Neuroendocrine tumor	12 (15.0%)	3(15%)	9(15%)	
Pancreatic adenocarcinoma	60 (75.0%)	15 (75.0%)	45 (75.0%)	
Neoadjuvant chemotherapy	8 (10.0%)	4 (20.0%)	4 (6.7%)	0.085

6.3. Results

6.3.1. Intraoperative Outcome

The majority of patients (70.0%) underwent pylorus-preserving pancreaticoduodenectomy (PPPD). The mean operation time was 285.3 ± 72.7 minutes (Table 2). Extended organ resection was performed in 40% of patients in both groups. Although not statistically significant, a higher percentage of patients in the ARAS group required vascular resection compared to the control group (40.0% vs. 23.3%, p = 0.107). None of the patients in the control group had arterial resection, whereas 15.0% (n = 3) of patients in the ARAS group underwent this procedure due to tumor infiltration (p= 0.002). Nevertheless, the operation time was significantly shorter in the ARAS group compared to the control group (245.9 ± 50.6 minutes vs. 298.5 ± 74.5 minutes, p = 0.004). While not statistically significant, it is clinically notable that the ARAS group had lower rates of excessive intraoperative bleeding (5.0% vs. 21.7%, p = 0.089) and the need for intraoperative RBC transfusion (0.0% vs. 13.3%, p = 0.085). Additionally, the ARAS group required significantly fewer RBC transfusions during surgery (0.0 ± 0.0 units vs. 0.5 ± 1.4 units, p = 0.014). None of the patients in the ARAS group had a positive resection margin (R1), whereas 20.0% of patients in the control group were diagnosed with an R1 resection margin (p = 0.045).

Table 6.3.: Intraoperative Data

. Bold value indicates all significant values with a two-sided P value less than 0.05. *Only patients with malignant tumors were analyzed (n = 72). PPPD indicates pylorus-preserving pancreaticoduodenectomy.

Variables	Total (n = 80)	ARAS (n = 20)	Control (n = 60)	*p*
Type of surgery				0.944
PPPD	56 (70.0%)	14 (70.0 %)		
Distal pancreatectomy	12 (15.0%)	3 (15.0%)	9 (15.0%)	
Total pancreatectomy	10 (12.5%)	1 (5.0%)	9 (15.0%)	
Appleby procedure	2 (2.5%)	2 (10.0%)	0 (0.0%)	
Extended organ resection	32 (40.0%)	8 (40.0%)	24 (40.0%)	0.999
Vascular resection	21 (26.3%)	8 (40.0%)	13 (21.7%)	0.107
Arterial	3 (3.8%)	3 (15%)	0 (0.0%)	**0.002**
Venous	19 (23.8%)	6 (30.0%)	13 (21.7%)	0.448
Excessive intraoperative bleeding (≥1000 mL)	14 (17.5%)	1 (5.0%)	13 (21.7%)	0.089
Intraoperative RBC transfusion				
Patient	8 (10.0%)	0 (0.0%)	8 (13.3%)	0.085
Unit	0.4 ± 1.3	0.0 ± 0.0	0.5 ± 1.4	**0.014**
Operation time (min)	285.3 ± 72.7	245.9 ± 50.6	298.5 ± 74.5	**0.004**
Positive resection Margin (R1)*	11 (15.3%)	0 (0.0%)	11 (20.0%)	**0.045**

6.3.2. Postoperative Outcome

As shown in Table 3, the ARAS group had a significantly shorter hospital stay, averaging four days less than the control group (13.8 ± 6.6 days vs. 17.9 ± 8.2 days, p = 0.046). There were no statistically significant differences between the groups in the rates of postoperative pancreatic fistula (p = 0.620) and delayed gastric emptying (p = 0.696). PPH occurred in 11.3% of patients (n = 9), with only 3.8% classified as grade C PPH. The rate of PPH was lower in the ARAS group (5.0% vs. 13.3%, p = 0.307), and none of the ARAS patients experienced grade B/C PPH. Major morbidity occurred in 17.5% of patients (n = 14), with a lower incidence in the ARAS group compared to the control group (10.0% vs. 20.0%, p = 0.308). The 30-day mortality rate was 6.3% (n = 5), with no significant difference between the groups (p = 0.790). Only one patient in the ARAS group died; this patient had an arterial anomaly, including an accessory right hepatic artery originating from the superior mesenteric artery, and underwent extended organ resection and portal vein reconstruction. The ARAS technique provided excellent vascular identification, resulting in a 292-minute operation time and 400 ml of blood loss during surgery. Unfortunately, the patient developed severe postoperative pneumonia and respiratory insufficiency and died eight days after the operation.

Table 6.4.: Postoperative Data. Bold value indicates all significant values with a two-sided p value less than 0.05. *Based on the ISGPS definition, 10 patients who underwent total pancreatectomy were excluded from this analysis. †Based on the ISGPS definition. ‡≥ Gade IIIb based on the Clavien-Dindo classification.

Variables	**Total (n = 80)**	**ARAS (n = 20)**	**Control (n = 60)**	***p***
CU stay (days)	3.4 ± 1.3	2.6 ± 5.2	3.7 ± 5.1	0.415
Hospitalization (days)	16.9 ± 8.0	13.8 ± 6.6	17.9 ± 8.2	**0.046**
Postoperative	11 (15.7%)	2 (10.6%)	9 (17.6%)	0.739
pancreatic fistula*				
Biochemical leak	4 (5.7%)	1 (5.3%)	3 (5.9%)	0.449
Grade B	4 (5.7%)	1 (5.3%)	3 (5.9%)	
Grade C	3 (4.3%)	0 (0.0%)	3 (5.9%)	
Delayed gastric emptying†	10 (12.5%)	2 (10.0%)	8 (13.3%)	0.696
Grade A	7 (8.8%)	1 (5.0%)	6 (10.0%)	0.756
Grade B	3 (3.8%)	1 (5.0%)	2 (3.3%)	
Postpancreatectomy hemorrhage†	9 (11.3%)	1 (5.0%)	8 (13.3%)	0.307
Grade A	2 (2.5%)	1 (5.0%)	1 (1.7%)	0.374
Grade B	4 (5.0%)	0 (0.0%)	4 (6.7%)	
Grade C	3 (3.8%)	0 (0.0%)	3 (5.0%)	
Major morbidity‡	14 (17.5%)	2 (10.0%)	12 (20.0%)	0.308
30-day mortality5 (6.3%)	1 (5.0%)	4 (6.7%)	0.790	

6.4. Discussion

In recent years, the introduction of more efficient neoadjuvant chemotherapeutic regimens for patients with borderline resectable pancreatic cancer has increased the hope that a greater number of patients may become eligible for potentially curative resections [79, 102]. However, many of these patients present with arterial involvement, necessitating radical pancreatic surgery, which is associated with a high risk of intraoperative complications [103, 181, 274]. Such complex procedures require a surgeon with substantial expertise in hepatopancreatobiliary surgery [204]. Despite this expertise, translating preoperative 2D CT imaging to the intraoperative setting remains a significant challenge. To enhance intraoperative vascular identification and improve surgical outcomes, we introduced ARAS, which has shown notable accuracy and effectiveness in ensuring precision and safety during vascular preparation and dissection (see Chapter 5). In this study, patients in the ARAS group experienced significantly shorter operation times compared to the control group, with an average reduction of approximately 50 minutes, a difference of notable clinical importance. It is important to highlight that the proportion of borderline resectable patients in the ARAS group was three times higher than in the control group,

and the rate of vascular resections was twice as high in the ARAS group. These factors have been shown to be associated with poorer perioperative and oncological outcomes [131, 181]. Nevertheless, although not statistically significant, the ARAS group experienced a lower rate of excessive intraoperative bleeding (5.0% vs. 21.7%). Additionally, none of the patients in the ARAS group required intraoperative RBC transfusions. These presented patients' outcomes lie within the range of previously published literature. Moreover, none of the patients in the ARAS group were diagnosed with a positive resection margin (R1). One of the most critical challenges in radical pancreatic surgery is the complete removal of the tumor and peripancreatic tissue while preserving the surrounding blood vessels [243]. Achieving this requires the total mesopancreas excision and precise preparation and skeletonization of vital peripancreatic vascular structures, such as the superior mesenteric vein and the right edge of the superior mesenteric artery, which is essential for a complete resection. Previous studies have demonstrated improved oncological outcomes in patients undergoing such radical procedures [42, 126, 243]. However, these procedures are typically associated with a higher incidence of intraoperative complications. ARAS, as a wearable AR-based navigation system, allows for the immediate registration of 3D models and accurate identification of vascular structures. It can also be continuously employed throughout the operation (especially the resection phase) if needed. We believe that utilizing these features of ARAS led to safer and more effective tumor removal, particularly from the critical area surrounding the right edge of the superior mesenteric artery, and thereby improved resection margin outcomes with a lower incidence of perioperative complications. Furthermore, PPH is the most serious complication following pancreatic resection. Although PPH is a rare complication in high-volume centers, mortality rates for patients with PPH have been reported to be as high as 88% [23]. Although not statistically significant, patients in the ARAS group experienced approximately threefold fewer instances of PPH compared to the control group. Only one patient in the ARAS group developed grade A PPH, who did not require any additional intervention. Additionally, the rate of major morbidity was twice as low in the ARAS group. As a result of these improved outcomes, the duration of hospital stay was significantly shorter in the ARAS group. Despite these promising results, there are certain limitations applied. Firstly, the sample size was small, and the study was not conducted as a randomized controlled trial. To minimize potential biases, we designed and conducted the ARAS study prospectively and performed a matched-pair analysis for data comparison.

Taken together, these clinical outcomes provide compelling evidence that ARAS offers meaningful benefits beyond the controlled environment of technical evaluation. The observed improvements in operative time, blood loss, transfusion rates, and resection margins, despite a higher proportion of complex cases, demonstrate the system's practical utility in high-stakes surgical settings. Most importantly, these findings validate the translational value of our engineering approach. By enhancing surgical precision and safety, ARAS has the potential to significantly improve patient outcomes in real-world clin-

ical practice.

Chapter 7

From Visual Clutter to Cognitive Support: LLMs and Context-Awareness

Contents

As AR systems become increasingly integrated into high-stakes environments like healthcare, their effectiveness depends not only on what they display but

also on how and when that information is presented. Without intelligent filtering and contextual awareness, even well-designed visual aids can contribute to cognitive overload, ultimately distracting rather than supporting users during critical tasks.

As outlined in Chapter 3, the selection of appropriate interaction modalities is a key determinant of user performance and system usability in XR applications for the medical domain. This becomes particularly crucial in surgical contexts, where spatial constraints, user-specific limitations, and elevated cognitive demands intersect, posing significant challenges to effective human–system interaction.

This chapter investigates the feasibility of integrating Large Language Models (LLMs) to enhance the usability and intelligence of AR-based surgical navigation systems. It demonstrates how LLMs can not only reduce users' cognitive load and improve task efficiency but also streamline system development by enabling dynamic orchestration of modular functionalities to support more context-aware and adaptive behavior.

The content of this chapter covers the following publication: .

1. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. "LLMs Enable Context-Aware Augmented Reality in Surgical Navigation." DIS25, Designing Interactive Systems Conference, 5-9 July 2025, Funchal, Madeira.

This work was awarded with "Honorable Mention Award" during DIS'25 conference in Funchal, Portugal where it was first presented.

7.1. Background

Wearable AR has become a technology with vast potential for surgical navigation systems, promising enhanced precision and real-time guidance for medical professionals. Despite its promising capabilities, the effective integration of AR in critical domains, such as open surgery, has faced delays and notable challenges compared to other professional fields [71, 81]. One of the challenges associated with this integration delay is the current limitations in usable interaction methods to control the system [47, 93, 225] that could be easily used and adapted to critical domains [52].

The interaction techniques used in wearable AR technologies might vary depending on the task and used modalities [108]. While many interaction modalities could be used for a range of tasks in non-critical medical domains, such as training and simulations[5, 135], the options for hygienic and sterilized surgical environments remain very restricted. Among different interaction and control modalities, hand, voice, and foot input became possible options for interactions in the surgical environment as they do not require any direct contact with foreign objects [106]. Despite the usability of these interaction methods, the environmental limitations during the surgical theater might not

always allow practical usage of all these input modalities [176, 266]. Moreover, the limited number of different gestures that could be performed using only hand or foot inputs might restrict the system features [108] when compared to voice.

Among alternative interaction methods, voice-controlled assistants using speech commands stand out as a viable option, offering a hands-free and potentially intuitive means of interacting with AR systems during surgical procedures [27, 106, 108, 176]. While impressive in their ability to respond to speech commands, often lack contextual understanding and adaptability [95]. Furthermore, relying solely on speech commands presents its own set of challenges, including the need to implement distinct keywords for each custom functionality, consequently increasing the complexity of the application for users [176]. In domains where simplicity, efficiency, and ease of use are mandatory, such as in critical surgical settings, introducing additional workloads or time-consuming processes may compromise the technology's adoption and effectiveness.

On the other hand, the employment of voice input beyond using only speech commands and more as a natural interaction to control and communicate with the system gathered extensive attention in the research. While the majority of applications focused on using voice in combination with other input modalities [118, 152, 192], the recent speech recognition algorithms and natural language processing provided mediums to use voice-based interaction as the sole input modality [253]. With the latest development of smart assistant systems and natural communication schemes through speech, voice input became popular, especially where other input modalities could not be used [112]. With the outbreak of LLM, new possibilities have emerged to use natural communication schemes for interaction with assistant systems. Mahmood et al. [165] presented a LLM-powered conversational voice assistant that could be used in different areas. The combination of speech input and LLMs could also be used to achieve smarter assistant systems to control the system and perform certain system functionalities, as it was demonstrated by Dong et al. [63].

Despite recent improvements in interaction methods for AR applications [192], and in voice-controlled assistant systems[70], a gap still exists in the feasibility of using these techniques for critical domains, such as surgical navigation systems [52, 225]. The unique challenges presented in such vital domains have not yet received sufficient attention.

7.2. Methodology

In this work, we focused on finding an optimal interaction method for ARAS (Chapter 5) specifically designed for open pancreatic surgery, addressing the challenges associated with the impracticality of common input modalities in confined surgical spaces. This led us to explore and evaluate voice control (VC) methods for a more practical and efficient user experience in this surgical context. To guide our research, we addressed the following research questions over five consecutive phases as depicted in Figure 7.1.

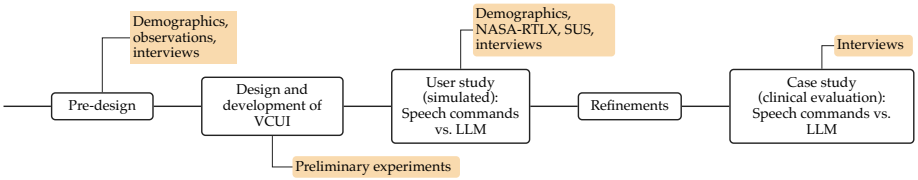


Figure 7.1.: Chart showing different phases of our work from pre-design till the case study involving a clinical trial and associated data collection for each phase.

Research Questions (RQs):

1. What are the user and field-specific requirements in terms of interaction methods for an AR-based surgical navigation system?

   Objective: To gather insights from surgeons to inform the design of the user interface and interaction with the AR system, ensuring it meets their practical requirements and enhances their ability to perform the surgery.

   Method: Observations, interviews with surgeons, along with demographics of participating surgeons.

2. How feasible is an LLM-based VUI for a surgical AR system, and how does it impact the user's cognitive load? How does this approach compare to previously tested VC methods, such as speech commands?

   Objective: To gather insights from surgeons about the usability, cognitive workload, and their assessment of the LLM-based VUI method and compare it to the previously employed approaches.

   Method: User study with surgeons in a simulated scenario involving surgically relevant system interaction tasks. Data collection using NASA_RTLX [105], and system usability scale (SUS) [30], and post-study interviews with surgeons, along with demographics of participating surgeons.

3. How feasible is the LLM-based VUI in the users' end setup during surgery compared to speech commands? What are the users' (surgeons') reflections?

   Objective: To gather insights from field surgeons about the usability of each VUI in an ecologically valid setup that involves highly stressful situations.

> ***Method:*** Case study involving the employment of both VUIs, each in a pancreatic surgery session, and conducting postoperative interviews with surgeons about the interaction method used to control the surgical AR system.

We began our investigation by exploring the user and domain-specific requirements for a VUI for ARAS by interviewing experts from the field. Consequently, we developed two non-conversational VUIs. Following the previous works on clinically tested voice interaction methods, our first VUI utilizes speech recognition and speech commands [60, 229] and serves as our baseline. The second VUI incorporates speech recognition, an LLM, and natural communication schemes to control the system. Internally, both VUIs have access to the same set of system functions of ARAS. In our first user study, we explored the usability of each VUI in a simulated environment and compared their performance in terms of added workload on surgeons while performing surgically relevant tasks, their usability, and conducted semi-structured interviews with participating surgeons. Finally, after proving the usability of both methods, we tested both VUIs in an ecologically valid environment during a clinical trial and conducted interviews with surgeons.

In this work, we solely focused on evaluating our proposed LLM-based VUI for ARAS in a time-critical domain, specifically in pancreatic surgery, and compared our approach to a conventional VUI using speech commands. The system design and evaluation are covered in Chapter 5.

7.3. Design and Development of LLM-Based System Interaction

Motivated by the interview outcomes from ARAS clinical trials and identified environmental and technical challenges, we further investigated the ways of optimizing the system interaction. We developed a new approach using LLM and a natural communication scheme and compared it with our previously tested VC methods for ARAS using speech commands. These speech commands included a total of 34 unique keywords that were assigned to trigger the system functionalities. The detailed on system design are given in Section 5.3. The list of used speech commands and their functionalities is also previously provided in Table 5.2 and Table 5.4.

Our new approach using LLMs included a system framework to automatically execute system functions upon user query stated via natural speech and speech recognition using the same Windows speech recognition service as the speech commands method [116]. We implemented a GPT communicator layer for Unity in C# to facilitate external API calls to chat-GPT from Unity and used GPT-3.5-turbo model to process the user request and return the system functions. Our LLM-based VUI framework consisted of four main components: a dynamic initial prompt generator, a dictation service, a response handler, and a GPT 3.5 model (Figure 7.2).

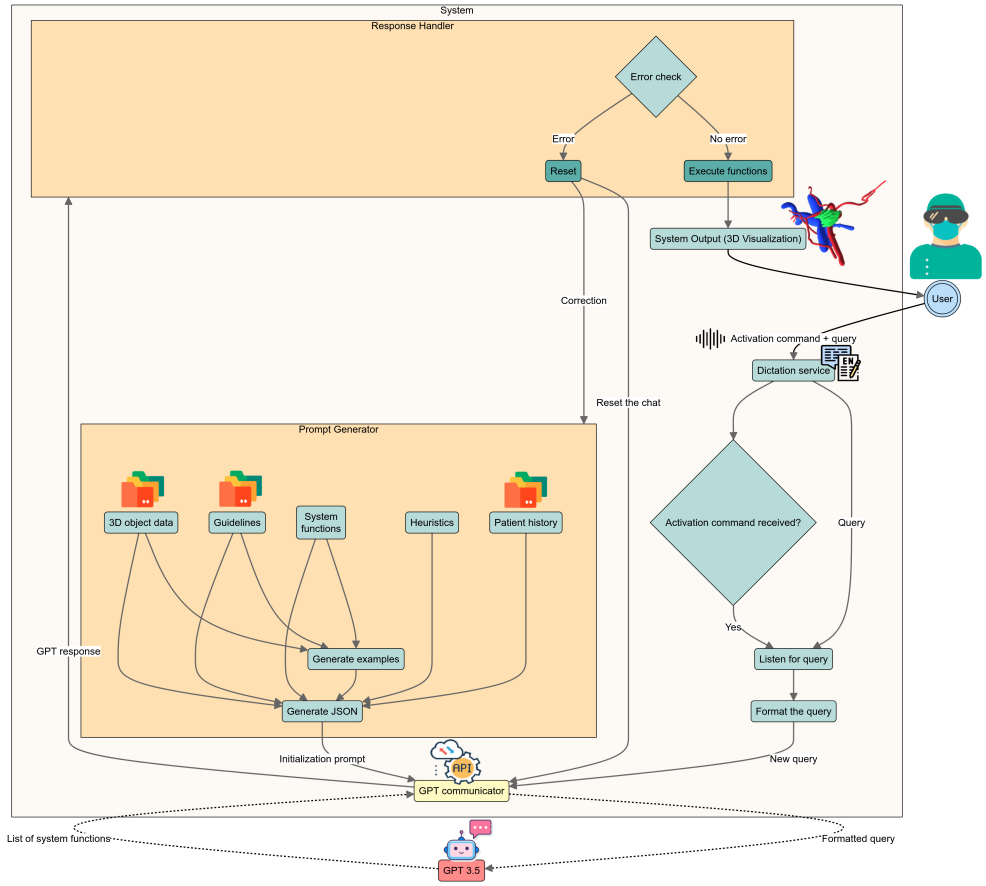


Figure 7.2.: Overview of LLM-based VC framework. The system begins with loading patient files and function descriptions to generate an initial prompt. The system functions are then called in response to the user's query via speech.

We developed an adaptive prompt generator to dynamically create initial prompts specific to each patient, aiming to provide patient-specific contextual information while defining the task for the LLM.

We used the patient's specific 3D model meshes to calculate the proximity and relational distance of each 3D object in the model, such as vessels and tumors, to other structures. We further provided the system with the patient diagnosis and surgical resection guidelines, along with a list of system functions, and heuristic examples (Figure 7.2).

To mitigate hallucinations by LLMs, as demonstrated in other related studies [91, 153, 276], we implemented an auto-repeat function that resends the initial prompt to the chat. Our preliminary experiments revealed that the accuracy of LLM outputs decreases as more user inputs are added, causing the model to lose the context of the initial prompt, which contains critical patient-specific information and the main task. This leads to increased hallucination in the generated responses.

To counter this effect, we developed a mechanism that periodically reminds the LLM of the task and relevant information by resending the initial prompt after each user input. Once the user's request is processed, this reminder prompt is automatically sent to the chat, ensuring that the model retains the original context without the user noticing.

As a safeguard for hallucinations that resulted in non-accurate response from LLM, we have implemented a reset function that would be used to correct the initial prompt and reset the chat. Like other system functionalities, the LLM could call upon this function based on the context of the user query. Users were informed about this functionality and instructed that if the system executed an incorrect or unexpected action, they could notify the LLM and specify the correct response for that situation. Apart from this reset function, no extra function or direct annotation to the study tasks (Section 7.4.1) was included in the initial prompt to avoid potential performance bias for the sake of the study.

As the GPT model only receives data in text format, we generated a JSON file containing all this information, along with the requested task to return the appropriate system functions and variables based on the given sentence. The JSON format used for the initial prompt is given in Listing 7.1.

```
1
2  Initial_prompt = {
3    ''description'': ''Depending on given sentences, Return
4    only appropriate method or methods from the executable
5    methods list without explanation.'',
6    ''executableMethods'': [
7      ''function_A(variables)'', ..., ''function_X(variables
         )''],
8    ''organTypes'': [
9      ''Organ_A'', ..., ''Organ_X''],
10   ''OrganCategories'': [
11     ''Category_A'', ..., ''Category_X''],
12   ''distanceData'': [
```

```
    `` Organ_A'': { `` Organ_A'': xx, ..., `` Organ_x'':
       xx},
    ...,
    `` Organ_x'': {`` Organ_A'': xx, ..., `` Organ_x'': xx
       }],
  ``guidelines'': [
    ``rule_A'':`` description of rule_A,
    ...,''
    `` rule_x'':`` description of rule_x'' ],
 ``sentencesAndResultsExamples'': [
    { ``sentence'': ``Show me all of the arteries'',
      ``result'': ``function_{xx} (variable_xx)''
    },
    ...,
    { ``sentence'': ``Show me the infiltrated vessels'',
      ``result'': {``function_{yy} (variable_y1,
         variable_y2)'',
      ..., ``function_{yy} (variable_y1, variable_y2)''
         }}]
}
```

Listing 7.1: Jason format used for sending initial prompt to GPT model.

After successful initialization, the user could send a request to the application through a voice query using a natural communication scheme. The dictation is activated using a single speech command called "Assistant" to avoid false queries. After activation of the dictation system, the system starts listening to the user's query. During preliminary testing, we observed that most of the requests from users to the LLM last around ten seconds. Therefore, we set the default listening time to ten seconds. However, if the ten seconds are exceeded and the user is still speaking, the system would wait for two seconds of silence before sending the query to the LLM. This way, we made sure that the system would at least listen to the user query for ten seconds while also allowing for longer queries. The transcribed user query would then be formatted to JSON and sent to the GPT model via the GPT communicator. Upon receiving the response from GPT, the response is first validated, and if there is no error in the received format or a request to reset the chat by LLM due to the user's correction, then associated system functions are executed and presented to the user.

If there is a call to reset among the received functions from LLM, then the system stores the user interaction example along with the correction that the user provided to LLM. The active chat would be terminated, another initial prompt would be generated using the recently added user example, and a new chat would be initiated.

In addition, similar to tooltips used in the speech commands approach, we designed a virtual panel (Figure 7.3) to provide a real-time transcription of the user's voice input and to provide visual feedback about the recognition of the user's query. This decision was made to show the user the recognized request and give the user a chance to correct it if it was detected incorrectly.

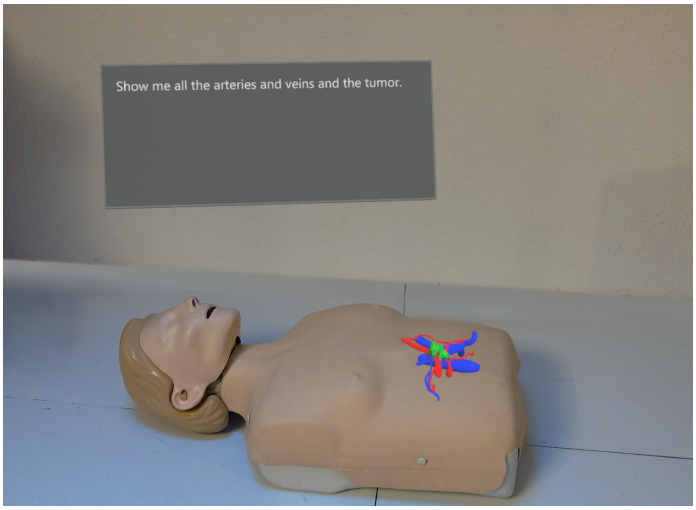


Figure 7.3.: A captured image from the AR surgical assistance system using LLM-based VUI. The manikin and transcription panel were used in the first study to simulate the visualization of the overlayed patient-specific 3D model during the surgical session.

7.4. User Study 1: Evaluation During Simulated Surgical Scenarios

We employed a mixed-method evaluation approach, integrating both qualitative and quantitative data analyses, to comprehensively assess the feasibility of our proposed LLM-based VUI and draw comparisons with the VUI method utilizing speech commands. In this study, we compared and evaluated both methods in a simulated lab environment, focusing on our two research questions (RQ2). We conducted a within-subject study, where all participants experienced both VUI (LLM and speech commands) for two different patient cases. To avoid potential bias, the order of the VUI and the patient case was counterbalanced.

7.4.1. Study Design

Aligned with ARAS's primary function, the study tasks focused on adjusting the visualization of 3D model segmentations to guide various phases of surgery. This approach aimed to test the VUIs within their relevant context in a controlled, simulated environment.

Even though the tested system's functionality was limited, this task design aimed to focus more on interaction with the system as a result of cognitively challenging tasks, which would be the normal case during the pancreatic surgery. The decision to trigger different system functions to visualize various combinations of structures depended on several factors, such as the relationship between structures and the tumor, patient history, surgical guidelines, and the current stage of the surgery. Any unnecessary virtual objects or structures visible in the AR view could confuse the surgeon or unnecessarily occlude the

view. It is worth noting that while it is possible to implement separate functions for certain predefined guidelines, often the decision of which structure combination to visualize is not definite, hard to implement, requires processing power on the operating device, and highly varies depending on each patient case and the stage of the operation.

Given the multifaceted nature of pancreatic surgery, which is divided into consecutive sessions for the preparation and resection of the vascular system and pancreas organ infiltrated with the tumor, we tailored the tasks to align with the interaction with the system during these distinct stages of each intraoperative session.

The task designs and their execution order were advised by experienced surgeons to simulate the progression of pancreatic surgery and the associated workload on surgeons. This approach aimed to replicate the decision-making process, considering the varying cognitive difficulty levels involved in identifying vital structures at different stages of the surgery.

The first two tasks targeted the commonly used first surgical approaches; therefore, the names of the structures to be visualized were given in the task description. Tasks 3 and 4 are performed in occasional situations based on the progress of the operation; however, due to the fix procedure performed in these tasks, the names of the essential structures to be visualized were also given in the task description. Tasks 5 and 6 were designed to address later stages of the operation, where the surgeon is required to make complex decisions on which structures need to be observed to guide a critical phase of the operation, such as the tumor resection phase. Therefore, in these last two tasks, the names of the structures to be visualized were neither given in the task description nor were annotated in the system, as the decision on which structure to be visualized is subjective to the surgeon and might vary. While task 5 focuses on visualization of the structures that are affected by the tumor, task 6 focuses on visualization of the structures that need to be removed along with the tumor, which is not always limited to those structures affected by the tumor. The task descriptions were as follows:

Tasks:

1. **Kocher maneuver:** During this task, participants were asked to only enable the visualization for the following structures: the tumor, inferior vena cava, and portal vein [263].
2. **Preparation of the hepatoduodenal ligament:** During this task, participants were asked to enable visualization for the following structures: portal vein, hepatic artery, and gastroduodenal artery [263].
3. **Uncinate-first approach:** During this task, participants were asked to enable visualization for the following structures: tumor, portal vein, and superior mesenteric artery [263].

4. **Artery-first approach:** During this task, participants were asked to enable visualization for the following structures: hepatic artery, gastroduodenal artery, celiac trunk, and superior mesenteric artery [263].

5. **Tumor infiltration:** Participants were asked to only enable only those structures that are infiltrated by the tumor. They were free to look at the CT images to decide or observe the 3D model, as it would be the case for the real surgery.

6. **Complex surgical decision:** During this task, the participants were asked to evaluate and make decisions about which structures should be resected (surgically cut or removed along with the tumor to secure patient safety) during the pancreatic resection and only enable the visualization of those structures.

During this study, we used a medical manikin to simulate the surgical scenario where the AR assistance system (Figure 7.3). We used the reconstructed 3D models of two real patients with complex pancreas tumor localization with vascular involvement to efficiently address all the above-mentioned tasks. Both study groups included both patient cases. The participants were asked to stand around the table where the manikin was placed and position themselves as the lead surgeon (Figure 5.12, surgeon number 1), as the main decisions during the surgery and the above-mentioned tasks are usually performed and are decided by the lead surgeon.

7.4.2. Measures and Data Analysis

We used quantitative and qualitative measures to evaluate different VUIs. TCT (measured in seconds) was recorded for each performed task. We also recorded the attempt count for the successful completion of each task.

We used the SUS [30] and NASA-RTLX [105] questionnaires to evaluate system usability and perceived cognitive workload after the completion of all tasks using each VC method. We concluded with a semi-structured interview with each participant to gather qualitative insights about their experiences with each method.

To analyze the quantitative data, we formulate the following hypotheses:

Hypotheses (Hs):

1. The LLM-based VC leads to lower task completion times.
2. The LLM-based VC leads to lower cognitive load as measured by NASA-RTLX.
3. The LLM-based VC has better usability as measured by SUS.

After confirming the normality of the data, we thus conducted a one-tailed

paired t-test to confirm or reject our hypotheses. For H3, we conducted a one-tailed Wilcoxon signed rank test instead, since data normality was violated.

All interviews were transcribed verbatim. Given the volume of the data, we followed the pragmatic approach to qualitative analysis as recommended by Blandford *et al.* [26]. Initially, two researchers analyzed 25% portion of the data. Following this, we created a preliminary coding framework through iterative discussions. The remainder of the interview data was then divided equally among the two researchers for coding. In a concluding discussion, we refined the coding framework further and identified overarching themes.

7.4.3. Procedure

The study procedure started with participants completing demographic questionnaires to provide essential background information. To mitigate order bias and potential learning effects, all participants engaged with both studied VUIs in a counterbalanced order with a different case for each VUI. Prior to task execution, participants were familiarized with the AR system and its functionalities using each VUI. Data recording was initiated after participants confirmed their ability to successfully interact with the system. Each participant systematically performed all experiment tasks for two sessions, each session using a different VUI and a different patient case. Task progression was subject to verbal confirmation from the participant regarding the accurate visualization of structures. Upon completion of the six tasks in each session, participants were prompted to fill out paper-printed questionnaires asking them to specifically answer the questions considering the experienced VUI but not the patient case. After completing both sessions, a brief semi-structured interview was conducted to gather insights about participants' opinions regarding each VUI. The whole study took approximately 40 minutes per participant. The study received the approval of the [removed for review] ethical committee board, ensuring that all aspects of the research adhered to established ethical guidelines.

7.4.4. Participants

Our study included nine volunteer experienced surgeons with a mean age of 42.44 (SD = 7.49) and 14.33 (SD = 7.35) years of average surgical experience. We compared the participant number with the participant number required for a usability evaluation. The number of participants in this study falls above the acceptable range of 4 ± 1 [37, 113], considering their high expertise in the field. The detailed characteristics of the participants are given in Table 7.1.

7.4.5. Results

Quantitative Measures: TCT, NASA-RTLX, SUS

The participants completed all six tasks using both methods: LLM, with a mean attempt count of 1.074 (SD = 0.328), and speech commands, with a mean attempt count of 1.370 (SD = 0.784) for successful completion. As

Table 7.1.: Participant Characteristics (N=9). Likert scale values range from 1 to 5, with 1 being the least and 5 being the most frequent or proficient. SD = Standard deviation

Characteristic	Mean	SD
Age (years)	42.44	7.49
Surgical Experience (years)	14.33	7.35
How many times used AR (1-5 Likert Scale)	3.00	1.50
How many times used LLM (1-5 Likert Scale)	2.78	1.20
English Proficiency (1-5 Likert Scale)	3.89	0.60

shown in Figure 7.4, individual one-tailed paired t-tests revealed that the LLM-based VC method yielded a significantly lower TCT for all tasks: Task 1 ($t(8) = -3.04$, $p < .01$), Task 2 ($t(8) = -2.33$, $p < .05$), Task 3 ($t(8) = -2.60$, $p < .05$), Task 4 ($t(8) = -2.56$, $p < .05$), Task 5 ($t(8) = -4.34$, $p < .01$), and Task 6 ($t(8) = -4.16$, $p < .01$). This result confirms H1.

The overall NASA-RTLX score was significantly lower for the LLM-based approach ($t(8) = -2.24$, $p < .05$), confirming H2. Specifically, it scored significantly lower for the subscales mental demand ($t(8) = -2.71$, $p < .05$), physical demand ($t(8) = -2.35$, $p < .05$), and effort ($t(8) = -2.36$, $p < .05$). We found no significantly lower score for LLM-based VUI for the other subscales, temporal demand, performance, and frustration. A visualization of this result can be found in Figure 7.5.

Using Bangor et al. [18] rating scale, the LLM-based VUI was classified as "excellent" with a SUS score of 87.78, and VUI using speech command was classified as "good" with and 79.17 SUS score. Despite the better performance of LLM, it did not yield a statistically significant higher SUS score. Thus, H3 could not be confirmed.

Interview Findings

Our analysis identified three main themes: User preferences and experience, context of use, and limitations and future improvements

User Preferences and Experience: Participants generally expressed a strong preference for the LLM VUI over traditional speech commands. They appreciated the flexibility and intuitiveness of the LLM, which allowed for more natural communication and seemed to reduce stress by accommodating various phrasings and intents without requiring specific keywords.

One surgeon reflected on the ability to articulate complex queries and receive accurate, contextually relevant information, which was particularly valued in high-stakes environments like operating rooms, saying:

> *"I personally think it's much less stressful to have such language support because I can say whatever I want, however I want to phrase it, and the system realizes what I want." (P1)*

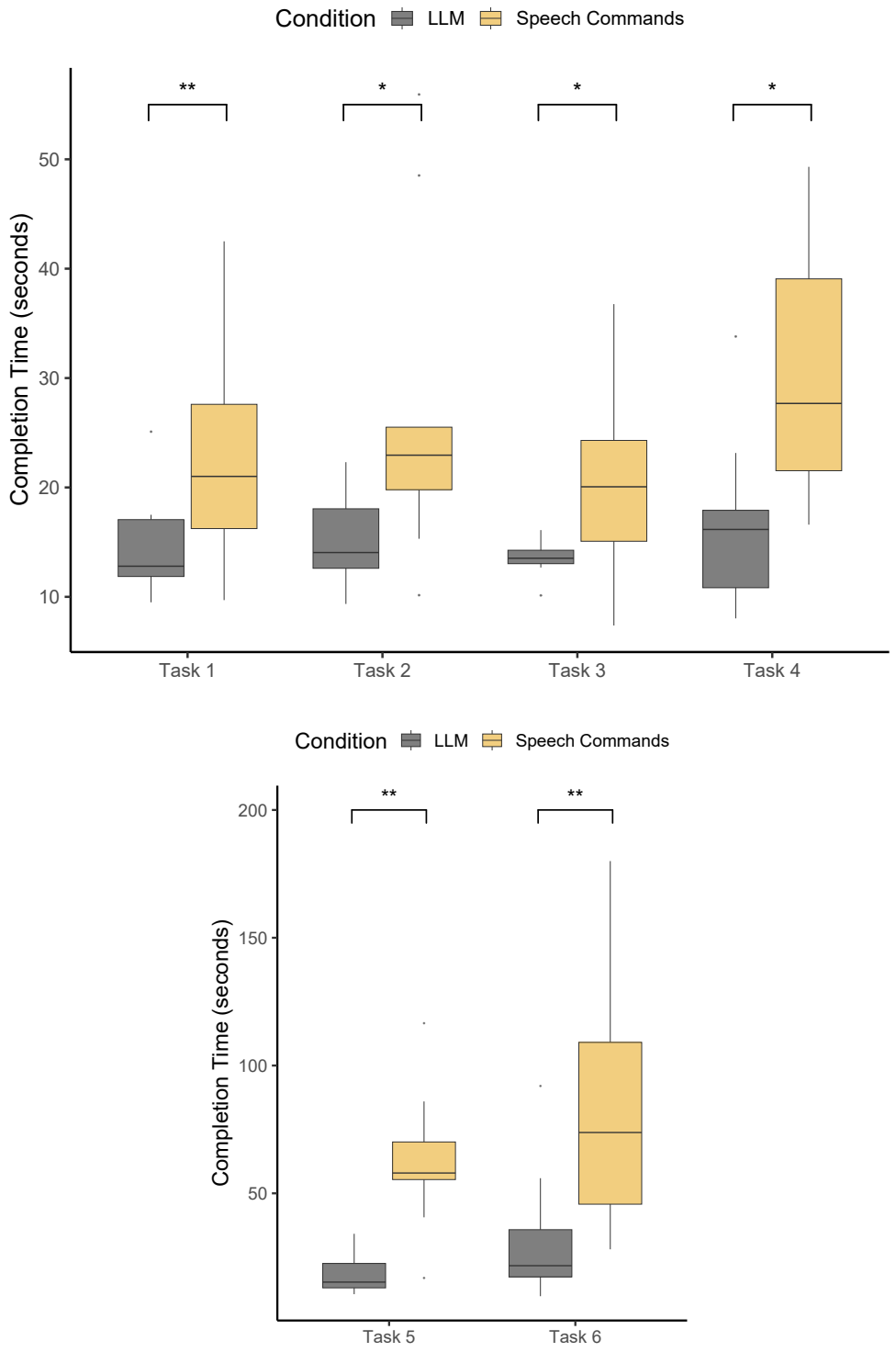


Figure 7.4.: Task completion times for each task given the different VC methods (LLM, Speech commands). LLM yielded significantly lower completion times for all tasks (marked with * for $p < .05$ and with ** for $p < .01$).

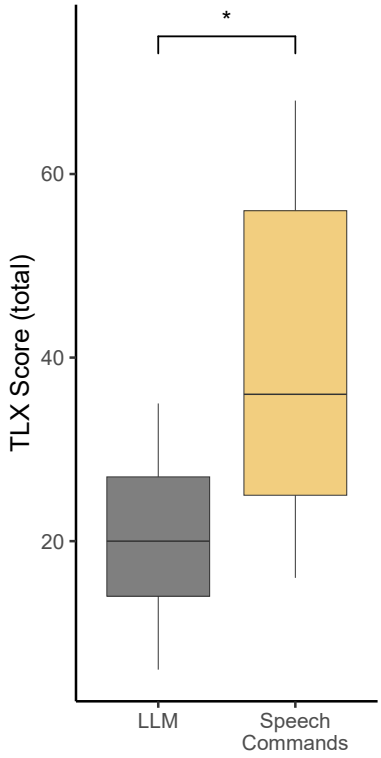


Figure 7.5.: NASA-RTLX score given both interaction methods (LLM, Speech commands). LLM yielded a significantly lower score (marked with *).

Furthermore, they emphasized that not only did the LLM-based VUI provided a more natural way of communicating with the system, but it also helped them to reduce the burden of thinking and making decisions about the requirements of the task by performing context-aware function calls. P2 pointed out this aspect, stating:

> *"For example, for tumor infiltrations, you have to first look at the tumor and which vascular system is infiltrated, then tell each time what to open or turn on [Speech Commands]. That's why I think it was much better with LLM, with a large language model, because I simply asked and that showed." (P2)*

Context of Use: The interviews revealed that, while both VUIs are usable, the specific context and situations in which the system is employed may highlight the unique potential of each method. While the LLM-based method could be highly beneficial in stressful and time-critical conditions, the speech command might be a better option for surgeries that do not have time-criticality. P8 and P9 highlighted this by the following statements:

> *"It depends on the time I have to use. So, for example, if you say there's an emergency, see, then I would prefer the AI [LLM]. In*

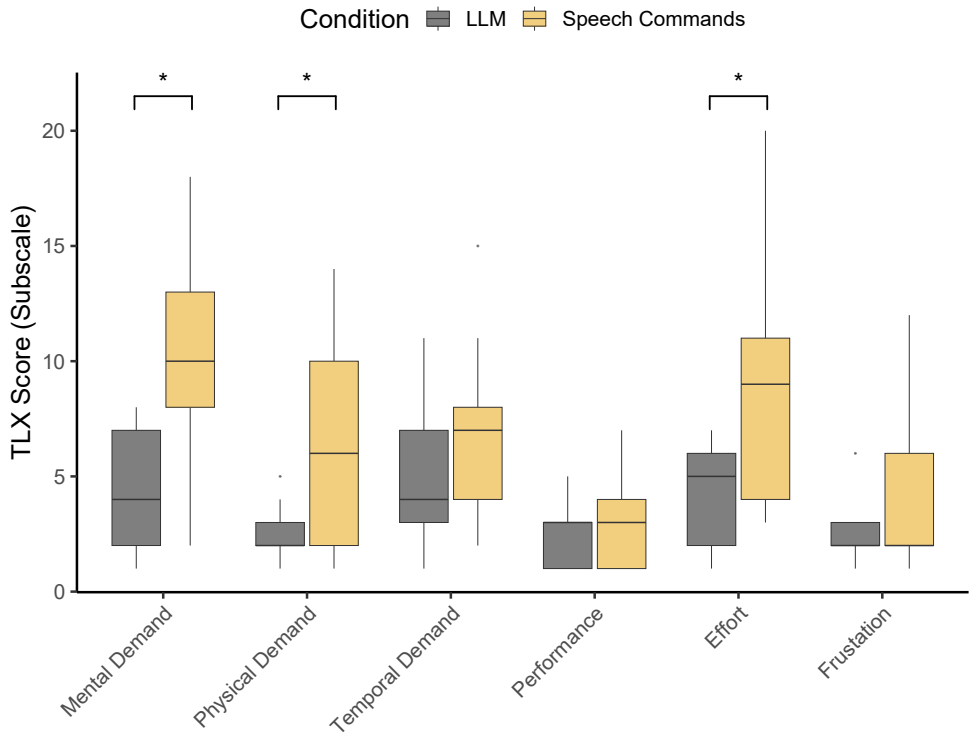


Figure 7.6.: Scores for individual NASA-RTLX subscales given both interaction methods (LLM, Speech commands). LLM yielded significantly lower scores for physical demand, mental demand, and effort (marked with *).

> *obesity surgery, I have time. I have no emergency. So I don't have a tumor, and I have more time. In the process of the operation, you can say remove this, remove that. So step by step [Speech Commands]." (P9)*

> *"If the patient bleeds, even a few seconds wait is already annoying (...) nevertheless, even with bleeding I find the second one better [LLM], because you can say directly what you want to see without thinking." (P8)*

Limitations and Future Improvements: The interviews revealed the importance of accurate dictation and robust speech recognition service for both VUIs. The visual feedback on the real-time transcription of the user query in the LLM-based method caused extra confusion, showing the potential misjudgment about the system's capability by the user. As users attempted to correct misinterpreted words upon observing incorrect transcriptions, the clarity of their requests diminished, leading to decreased performance in triggering the relevant functions by the LLM. Conversely, the LLM would typically mitigate such errors by interpreting them as typographical mistakes, thereby maintaining higher accuracy in understanding and responding to user commands. P4

pointed out this matter by raising attention to the system's capability being affected by misinterpreted words due to the different pronunciations, saying:

> *"The less I say, it's supposed to be easier for the system to understand me, right? I don't know. I was thinking that when I say too much, the system doesn't understand me because I said too much. With shorter words, you minimize the misunderstanding when someone pronounces it differently." (P4)*

The surgeons also raised concerns about the default listening time that was set in the LLM-based approach. P3 suggested adopting an approach where the initialization and ending of the listening time could be activated by some keywords to refrain from waiting if the query time is less or longer than the default listening time, saying:

> *"For example, you say assistant or something to begin. But can I also say end of sentence or so that I don't have to wait those couple of seconds." (P3)*

Additionally, participants mentioned that the LLM-based VUI required a clear statement of the request. Despite the ease of use, the interaction would be even easier over time as one would learn how to clearly phrase their request. P3 reflected on this saying:

> *"When I tried the second one [LLM], with the AI, I think it's even easier to use if you've done it before. Then you have the routine of what you have to say so that the device understands what I want, and then it's easier and quicker. I mean how I should formulate my question so that the system understands me and shows the result that I want." (P3)*

7.4.6. Implications

Our study comparing LLM-based VUI to a VUI using speech commands has revealed implications for the integration of such technologies into surgical settings. These implications highlight the potential benefits and necessary refinements for practical application:

Potential superiority of LLM-based approach in critical surgical moments: The LLM-based approach demonstrated advantages, particularly when a decision-making situation was involved. It exhibited significantly reduced execution times across various tasks. As tasks increased in cognitive demand, particularly in Task 5 and Task 6, the disparity in execution times became more pronounced, reflecting the challenge of mental workload and decision-making when using speech commands. Moreover, assessments of cognitive load indicated a lower mental demand with the LLM-based approach. This convergence suggests that LLM-based VUI could offer a superior option in real surgical environments, where timely decisions are required during constrained time frames.

Need for system refinements prior to real-surgery evaluation: However, our findings also illuminate areas necessitating refinement before practical deployment in surgical settings. The real-time transcription panel introduced confusion as users attempted to rectify misinterpreted words, compromising the clarity of the sentence context. Additionally, while LLM-based VUI facilitated quicker task completion, further enhancements in dictation service are essential to mitigate any remaining delays and optimize task execution times.

7.5. Case Study: Evaluation During Actual Surgery

Following the proven usability of both VUI systems in our initial user study (Section 7.4), this study aimed to further evaluate these VUIs in a real surgical setup, addressing our third research question (RQ3). This phase sought to confirm our findings under actual surgical conditions, which can differ significantly from laboratory environments due to factors such as higher stress levels and time constraints. Building upon our findings from our first user study (Section 7.4.5), we first performed refinements (Figure 7.1, Refinement) to our approach and later evaluated each VUI during a pancreatic tumor removal surgery (Figure 7.7).

We performed the following refinements to our LLM-based VUI: The transcription panel providing real-time feedback to visualize the transcription of the voice query panel was removed, as it was observed to cause more confusion. Users attempting to correct what they perceived as misinterpreted words during their query can diminish the efficiency of the LLM method, as the context of the sentence may become unclear. Instead, we used conversational audio feedback similar to those commercially available conversational assistants, such as Siri [1]. We used sound saying "OK" to indicate receiving the user query, and "Please state your request differently" when the LLM response did not yield any of the defined system functions. Furthermore, we adapted the listening time after activation of the dictation service. The user request would automatically send to the LLM model after one and a half seconds of silence without waiting for any further default time.

After performing refinements, we deployed our previously developed AR Assistance system designed to visualize the 3D model of the patient during the surgery, with the capability of both VUIs.

7.5.1. Study Design

To evaluate our LLM-based VUI and compare its outcomes, such as TCT and cognitive load, with the speech command during actual surgery, we exclusively employed qualitative measures and conducted post-operation interviews. This decision was driven by the inherent variability in each patient case, which might inevitably affect cognitive load measurements due to the unique nature

[1] Siri is Apple's voice-activated virtual assistant, available on iOS devices such as iPhones and iPads.

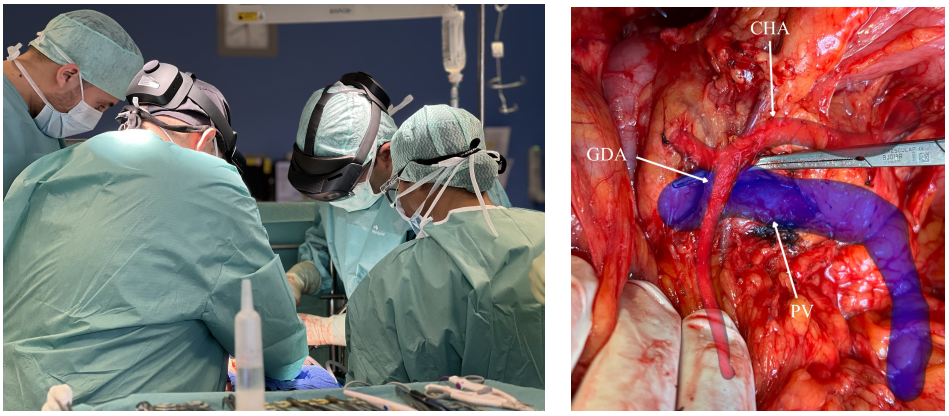


Figure 7.7.: Pancreatic Surgery session. The right picture shows a snippet from the application view captured from a surgeon's device. GDA: Gastroduodenal Artery, PV: Portal Vein, CHA: Common Hepatic Artery

of each surgical procedure. Similarly, task execution time would be influenced by the specifics of the surgery being performed. Consequently, a direct comparison between the two methods across different surgeries would not yield meaningful results. Thus, we chose to gather insights through interviews and observations, with an observer researcher participating in the surgery sessions, making notes and observations about user interaction with the system. This approach allowed us to assess the impacts of each VUI, enabling us to collect qualitative data that could inform future improvements and implementations.

7.5.2. Procedure

To validate our approach in an ecologically valid environment, we conducted clinical trials involving the intraoperative evaluation in patients with underlying (borderline) resectable pancreatic tumors who required various types of pancreatic resection. The trials took place at [removed for anonymized review], certified to perform pancreatic tumor resection surgeries.

The study protocol received approval from the Ethics Committee of the Medical Association of Saarland under registration number: (registration number: 159/23).

The protocol of our study was also registered at ClinicalTrials.gov under the registration number: NCT06208579.

Patients provided informed consent prior to surgery. All participating surgeons were fully briefed on the experimental nature of the method and the device used. They voluntarily agreed to use the system during the surgeries, assuming full responsibility for its operation and the outcomes. The surgeons were also informed that they could discontinue the use of the system at any point if necessary, without obligation.

To avoid first-time use bias, two of the surgeons (P1, P2) who participated in study 1 participated in this study. Each surgery began with two surgeons equipped with our designed wearable assistance system with both VC method

capabilities. In each session, surgeons were asked to use only one of the VC methods. However, they always had the choice to use the other method if it was essential for the course of the surgery. We conducted interviews with surgeons after each surgery session about their experience with each VUI.

7.5.3. Results

No technical difficulties regarding the VUIs were observed during both sessions, and both surgeons used the system with the assigned VUI throughout the surgery.

Interviews with two surgeons who experimented with both VUI across two surgical procedures proved the feasibility of our LLM-based VUI in a real surgical environment in line with findings from our first study.

The LLM's capability to discern user context and analyze patient data facilitated the surgeons in adjusting the visualization of patient 3D models according to the tumor's proximity more efficiently, specifically in the initial preparation phase of the operation. This benefit was encapsulated by a participating surgeon, who remarked:

> *"Today's patient had an anatomical anomaly, so we had to be more careful in identifying the vessels during preparation, so we don't damage them because they were so close to the tumor. So we had to change the visualization a lot. At that moment actually the LLM was a big help because it saved us a lot of time." (S2)*

S1 also reflected on this topic, saying:

> *"I think the biggest difference between the two [VUIs] was during the initial phase of the operation, where we usually use the system more to identify the vessels. But when the vessels are identified and already visible, we don't interact with the system much." (S1)*

Unlike the speech command method, which requires precise pronunciation of predefined keywords, the LLM system maintained a more natural communication flow. This aspect was profoundly appreciated, as S2 shared:

> *"speech commands also worked fine, but the issue with the speech commands is sometimes when people are talking around the table, you have to say a word 100 times until the system detects it. LLM is more forgiving if that's a correct word to use." (S2)*

Additionally, S1 reported on the benefits of using natural communication to control the system and also sharing information with other staff around the table. S1 reported:

> *"When I say a single word, usually other staff surgeons don't know what I am doing because they don't see what I see in the device. But when I talk to the system the way I talk normally, then*

> *they know, ok, now I am trying to see where some vessels are when I say, for example, show me the vessels near the tumor, or like I want to see the mesenteric artery." (S1)*

S2 also highlighted the benefits of the performed refinements, noting the reduced confusion from removing the transcription panel and improved system response times:

> *"This time with the LLM system, we didn't have to wait much for the system to react, so it was way better and less annoying. Also, I think removing the panel was a good decision, as I didn't see what the system understands so I didn't worry much about correcting my request, and the system worked even better." (S2)*

7.6. Discussion

The introduction of AR-based surgical assistance systems has significantly transformed surgical practices, offering an enhanced level of precision and support. As these technologies evolve, the choice of interaction method to control the system becomes a pivotal consideration. Our study introduces a novel VUI for surgical ARAS using speech recognition and LLM and conducts a comparative analysis with the conventional VUI using speech recognition and speech commands, focusing on enhancing operational efficiency and user experience in the critical context of surgery. Importantly, we tested both methods in controlled laboratory settings and real surgical environments, offering a robust evaluation of their practical application and performance. This dual-context approach allowed us to gather comprehensive insights into the efficiency, cognitive load, user preferences, limitations, and situational applicability of each VC method.

7.6.1. Efficiency and Cognitive Load: LLMs versus Speech Commands

The SUS scores, along with the successful implementation of both VC methods in simulated and real surgical environments, demonstrate their usability and confirm their applicability in critical medical settings. However, the distinction in performance, especially in time-sensitive scenarios like the initial phases of surgical intervention, which is the most mentally demanding phase, underscores the criticality of choosing the right control and interaction method.

The use of LLMs significantly outperformed traditional speech commands in TCT. This efficiency is attributable to the LLMs' ability to generate context-aware outputs and execute multiple functions simultaneously to achieve a certain undefined functionality, a quality unattainable with keyword-specific speech commands without implementing further keywords to perform this task. As functionalities expand, the speech command method suffers from scalability issues, requiring an ever-increasing list of keywords. Conversely, LLMs streamline this process, enabling parallel function execution based on user requests without necessitating an extensive set of unique commands.

A more intriguing aspect of LLMs lies not just in determining which function to call based on user requests, but also in acting as an intelligent assistant and generating outputs that normally require a complex decision-making process. This was particularly evident when LLM successfully generated the correct output to call certain system functionalities even when the function name or its specific purpose was not directly mentioned in the user's query. The capability of LLMs to generate context-aware outputs using all available information represents a significant advancement towards truly intelligent user interfaces and assistance systems.

A detailed analysis of user interaction logs with the LLM revealed its ability to successfully make decisions in numerous instances where users would otherwise have had to decide for themselves. This difference in performance compared to speech commands was particularly evident. Despite the system functionalities being simple and identical in both cases, the reasons for employing these functionalities were often complex. With speech commands, the user (a surgeon) needed to decide and then instruct the system to make specific changes in visualizations using keywords. In contrast, the LLM-based system handled the decision-making process.

For example, in task 6, P2 asked the LLM, "Can you show me what should be resected?" Despite no specific information or annotation regarding the task description or the structures to be enabled in this context, the LLM correctly decided to display the tumor and the infiltrated veins and arteries within the resection margins by invoking multiple system functions simultaneously. This decision-making process is highly complex, relying on factors such as patient history, tumor position, and surgical guidelines regarding resection margins.

This feature, demonstrated by the LLM not only reduced task completion times but also significantly diminished cognitive load. This was also evidenced by the lower scores in the NASA-RTLX, indicating a more intuitive and less burdensome interaction for the user, something unachievable with speech command methods without pre-implementing more complex functions into the system.

In applications such as surgical navigation systems, where system interaction is part of a decision-making process contingent on the patient case and scenario, pre-implementing an all-encompassing solution is very challenging. On the other hand, other specialized voice assistance methods, such as those using machine learning [95], require training the system with specific user terminology for different scenarios. Additionally, they come with significant processing costs that affect the performance of wearable devices. Here, LLMs can provide significant benefits, offering a more adaptive, context-aware, and efficient approach to managing complex tasks, even though the system itself remains simple regarding the functionalities.

By analyzing patient-specific data and examples given in the dynamically generated initial prompt, LLMs can offer custom recommendations, enhancing the support system's utility in high-pressure situations. This capability to interpret context and call relevant function combinations offers surgeons a richer, more contextual understanding of the patient's data, including visu-

alization of details in the 3D models, an attribution that with conventional voice assistant systems using speech commands cannot be achieved [95].

Despite the apparent advantages of the LLM method in facilitating quicker, multi-functional requests, our study also highlighted a perception mismatch among some participants. They perceived speech command execution as a faster method for task completion, despite objective evidence showing the LLM method reduced TCTs. This discrepancy may be linked to the system's default listening time (minimum ten seconds or wait for 2 seconds of silence if longer than ten seconds) following the user query. It suggests the necessity for an adaptive approach in managing the activation and deactivation of the system's listening duration for the LLM-based method to ensure the receipt of full user queries without a long wait. As a shorter listening period could prematurely send incomplete queries to the LLM, while a longer period might unnecessarily delay the system's response.

7.6.2. Pros and Cons: Balancing Control and Transparency

In our study, we found out that speech commands, with their direct and deterministic nature, afford users a clear understanding of cause and effect. This transparency in interaction fosters a sense of reliability and control, essential in high-stakes environments like surgery. However, this method's scalability and flexibility are constrained by the need to predefine every command, which can limit the system's responsiveness to complex or unforeseen requests.

On the other hand, LLMs represent a paradigm shift towards more fluid, conversational interactions. By understanding and processing natural language, these models offer a dynamic and flexible interface that can interpret a broad spectrum of user requests. However, this sophistication comes with a degree of opacity. The "black box" nature of LLMs can obscure the pathway from request to action, potentially undermining user confidence if the system's reasoning and decision-making processes are not sufficiently transparent.

Our study revealed that the LLM system could effectively compensate for errors in the dictation and speech recognition system. Unlike speech commands, which necessitate precise pronunciation, the LLM system can infer the user's intent by analyzing the context of the query rather than focusing on individual words. This capability significantly enhances the system's flexibility and user-friendliness.

However, it became evident that providing users with real-time visual feedback of transcription could inadvertently lead to misjudgments about the system's capabilities. Users attempting to correct what they perceive as misinterpreted words during their query can diminish the LLM method's efficiency, as the context of the sentence may become obscured. This observation underscores the critical need for designing user feedback mechanisms that do not compromise the clarity of communication or the efficiency of the system control method.

Furthermore, mitigating these challenges necessitates clear communication about the system's operational boundaries and capabilities. Users need to understand not just how to interact with the system, but also the underlying

principles guiding its responses. This understanding is crucial for formulating effective requests, especially with LLMs, where the context and specificity of language can dramatically influence outcomes. Training and educational programs play a pivotal role in this regard, equipping users with the knowledge to navigate the system's complexities and leverage its full potential.

We believe that a hybrid approach, integrating both speech commands and LLM capabilities, emerges as a promising solution to balance control with transparency. By allowing users to switch between modes based on the task's complexity or urgency, such a system combines the directness of speech commands with the adaptability of LLMs. For routine tasks or when precision is paramount, predefined speech commands could offer the most efficient pathway. Conversely, for tasks that require a time-consuming decision-making process or when additional context is required, the LLM recommendations could provide a faster solution.

Implementing a hybrid model also entails designing interfaces that intuitively signal which mode is in operation, thereby maintaining user awareness and trust. Visual or auditory cues could indicate the system's current state, whether executing a direct command or processing a more complex LLM-based request. Moreover, offering users the ability to override or specify the control mode empowers them to use the system's capabilities according to their immediate needs and preferences. By carefully navigating the trade-offs between control and transparency, and by fostering an environment of continuous learning and adaptation, we can develop systems that not only enhance surgical outcomes but also align with the users' operational and cognitive needs.

7.6.3. Ethical Considerations

Maintaining ethical standards is crucial for preserving trust in medical research and innovation. By adhering to ethical guidelines, researchers and practitioners demonstrate their commitment to prioritizing patient safety and well-being over technological advancements.

In this study, we emphasized ethical adherence throughout all stages. We ensured that the introduction of the AR system did not compromise the safety of patients or surgeons, nor did it undermine the integrity of the surgical process.

To achieve this, we initiated the study only after obtaining full approval from the relevant medical ethics review board. All participants, including surgeons and patients, were thoroughly informed about the study, with their participation contingent upon a clear explanation and the collection of informed consent. Patients were made aware of the potential risks, benefits, and alternatives to ensure their participation was both voluntary and fully informed.

We also ensured that neither the AR system nor the VC method used replaced the surgeon's judgment, maintaining human decision-making in surgical procedures, and surgeons retained complete control over the system, consistent with the Fundamental Principles of Ethics [260].

Our key takeaways from this study regarding ethical considerations for future studies are as follows: The system should function solely as a supplementary tool to assist the surgeon without replacing the surgeon's expertise. The surgeon must retain ultimate decision-making authority, ensuring patient safety and adapting to the unique aspects of each case. Human intuition and experience should remain the final safeguard in surgical procedures.

7.6.4. Limitations and Future Work

Even though the results of this study are promising steps towards using LLMs not only as a VUI but also as intelligent assistants in the medical domain, our findings are limited to the specific functionalities of ARAS we used in this study. While these functionalities are integral to such applications, a broader understanding of LLM capabilities requires further research. This should involve more complex system functionalities and tasks to fully explore and validate the potential of LLMs in diverse and demanding scenarios. In our future work, we intend to broaden the scope of our investigation into the capabilities and opportunities presented by LLMs in surgical assistance systems beyond our current application as a function caller and VC method. By leveraging the advanced natural language understanding and processing capabilities of LLMs, we hope to uncover new ways in which these models can contribute to the enhancement of surgical outcomes, efficiency, and safety.

7.7. Conclusion

Our comparative study of two VUIs within an AR-based surgical assistance system highlights the distinct advantages and considerations associated with speech commands and LLM. We found that the LLM-based VUI offered significant improvements in operational efficiency and reduced the cognitive load of users by allowing for natural, conversational interactions and the ability to generate context-aware system behavior by executing multiple functions concurrently. However, the choice between LLMs and speech commands is not clear-cut, despite a higher preference towards LLMs; user preferences may vary based on perceived control, transparency, and the context in which the system is employed. While speech commands provide a sense of direct control and transparency, LLMs require clear instructions to function optimally, which can sometimes challenge users. The idea of a hybrid model emerges as a promising solution, aiming to combine the strengths of both approaches to cater to a broader range of needs and situations in surgical settings. Looking ahead, we plan to expand our exploration into the potential of LLMs as conversational assistants that not only could control the system but could participate more in the decision-making process, further enhancing the capabilities of surgical assistance systems. This study lays the groundwork for future advancements in surgical technology, emphasizing the importance of the involvement of end-users during design and evaluation and the need for systems that balance efficiency, cognitive ease, and adaptability to the fast-paced, complex nature

of surgical environments.

Chapter 8

Is Technical Excellence Enough? Toward Ethical, Usable, and Clinically Viable AR Systems

Contents

While the previous chapters concentrated on the technical intricacies of designing reliable AR systems that improve skill acquisition and provide cognitive support, this chapter takes a step beyond the technicalities. It shifts focus toward the broader, multifaceted considerations essential for designing and developing systems within the high-stakes context of the medical domain. Drawing on our extensive research experience, careful analysis of qualitative

data gathered from end-users, and insights from the broader fields of Human-Computer Interaction (HCI), this chapter explores the topic through four distinct lenses: interdisciplinary collaboration and the imperative of co-Creation, human-centered design and experience, clinical environment and workflow integration, clinical environment and workflow integration, and generalizability and domain-specific constraints. Together, these perspectives aim to offer a more holistic understanding of what it truly means to create effective and trustworthy AR solutions in critical healthcare environments.

Parts of this chapter were covered in the following publications:

1. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, and Jakob Karolus. "From Concept to Clinic: Multidisciplinary Design, Development, and Clinical Validation of Augmented Reality-Assisted Open Pancreatic Surgery." In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1-24. 2025.

2. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, and Jakob Karolus. "Breaking the Dogma: Misconceptions, Challenges, and Lessons Learned from AI and AR-Assisted Pancreatic Surgery" Proceedings of CHI '25 Workshop on Envisioning the Future of Interactive Health.

8.1. Human-Centered Design and Experience

In high-stakes medical contexts, where the cost of error can be severe, the usability and cognitive compatibility of technology are just as critical as its technical performance. AR systems have the potential to support clinical decision-making, enhance training, and reduce human error, but only if they are designed with a deep understanding of the people who will use them. Human-centered design (HCD) emphasizes designing with and for users, recognizing their workflows, mental models, stress factors, and practical limitations. In this section, we explore the importance of understanding user profiles, usability principles, cognitive load, trust dynamics, and the need for continuous learning support in AR design for the medical domain.

8.1.1. Understanding the User Profile

Designing AR systems for healthcare requires more than general usability heuristics; it demands a nuanced understanding of specific user profiles. A surgical resident, an ER nurse, and a radiologist all have vastly different contexts, cognitive demands, and expectations. These distinctions influence not only how users interpret augmented information but also how they respond under pressure.

HCI literature underscores the importance of situated actions, decisions and behaviors shaped by the context in which they occur [244]. In clinical settings,

these actions are time-sensitive, collaborative, cognitively challenging, and often mediated by institutional routines. Neglecting these contextual details can lead to AR systems that are technically functional but practically unusable.

One of the main defining factors of user profiles is the cognitive load they regularly endure. Medical professionals operate under intense cognitive demands. AR systems designed for that introduce excessive visual information, require multitasking, or demand constant attention-switching risk overwhelming users rather than aiding them.

Moreover, understanding user profiles must go beyond a broad definition of their professional roles. Even within the same user category, expectations and information needs can vary based on experience level, responsibilities, and moment-to-moment task demands. For instance, what a lead surgeon considers essential information for a surgical navigation system may differ substantially from what a junior assistant deems most valuable. This was clearly reflected in interviews conducted with surgeons who participated in the same procedures but played different roles.

One senior surgeon emphasized a need for targeted, minimal information to support decision-making in high-stakes moments:

> *"Is there a variation of the anatomy? Is there a special situation about this patient? And where actually in the operation am I? Am I near the vessels? So then I know now I have to take it slow and really take my time to dig out the vessels. And don't make a quick cut or something like that. I think that is really enough. I don't need to see everything all the time."*

In contrast, a junior surgeon highlighted how the AR system offered valuable guidance by making the spatial structure of the operation more legible:

> *"When I participated in this type of operation for the first couple of times, I never understood what the chief surgeon was doing, because I didn't know where he was going in, I didn't know which vessels he wanted to focus... With this system (ARAS) and visualizing all of the 3D models (of vessels) during operation, I think I understood this operation better as I could clearly see where the vessels are."*

These examples illustrate the importance of designing systems that are sensitive to user role, experience level, and task-specific needs. Techniques such as contextual inquiry, task analysis, and persona development are essential in early design stages to align AR functionality with the cognitive workflows and expectations of different users [193]. Neglecting these differences increases the risk of cognitive dissonance, underutilization, or outright rejection of the system.

8.1.2. Usability and System Design

Usability remains a cornerstone of system adoption and performance. According to Norman [193], systems must afford intuitive interactions, visible

affordances, and clear feedback. In AR interfaces, this becomes more complex due to spatial interaction, gesture inputs, and layering of digital elements onto real-world environments.

Endsley's theory of situation awareness is highly relevant here. She argues that effective decision-making relies on maintaining an accurate perception of the current environment, understanding its dynamics, and anticipating future states [74]. AR has the potential to enhance this awareness, if it delivers timely, relevant, and non-intrusive information. Otherwise, it becomes a cognitive burden and impairs performance, particularly during emergencies.

Effective system design in AR must therefore go beyond visual fidelity to include ergonomics, gesture consistency, and contextual clarity. These factors help reduce user friction and improve performance under real-world pressures.

A key design principle observed during testing sessions for ARAS with surgeons, and consistently emphasized across feedback interviews, was the philosophy of **"less is more."** This principle should guide multiple layers of system design: from the setup process to the user interface, interaction, and especially to information visualization strategies.

The information visualization must be minimalistic, highlighting only the most critical data relevant to the task at hand. Overloading the user with unnecessary or static data increases cognitive burden and distracts from decision-making. Instead, a progressive disclosure model should be used, offering details on demand when needed [235]. This approach aligns well with the dynamic and focused nature of the critical medical domain.

Additionally, the system initialization and setup should be as straightforward as possible. For instance, in busy operating rooms, there is minimal tolerance for cumbersome configuration or system calibration. A streamlined system initialization reduces preparation time, technical friction, and resistance from clinical staff.

These requirements were clearly encapsulated by a surgeon, who emphasized the importance of fast system initialization requirements, easy interaction, and details-on-demand in time-critical procedures.

> *"I think ideally the system should just run and show only the positions of the vessels and tumor once you run it without a need to do much, so we can just quickly check before starting vascular preparation or tumor resection. You don't have the time to waste on having a menu and choosing the options during the pancreas surgeries. It would be nice to know that you have the options to control what you see if you need it, but without spending too much time on it."*

However, while minimizing complexity, it is equally important to maintain a balance between simplicity and control. Over-simplification can undermine transparency and user trust. Users must understand why the system displays or hides specific information; otherwise, they may develop under-reliance or over-reliance on the system, both of which can be harmful in critical scenarios.

Trust is shaped not only by the technical accuracy of the system but also by interface cues, feedback mechanisms, and the perceived controllability of the system. Systems that allow users to validate, override, or explore underlying logic foster more calibrated and responsible trust. At the same time, designers must avoid overwhelming users with excessive justifications or options, which could slow down decision-making or cause frustration.

8.1.3. Learning Curves and Training Support

Even intuitive AR systems require training, especially when they introduce fundamentally new ways of interacting with digital content. In the context of AR, users are often required to manipulate and interpret spatial data, perform mid-air gestures, or engage with voice or gaze inputs, interaction modes that go beyond the flat, screen-based paradigms most users are familiar with.

The learning curve varies by user group and system complexity, but neglecting structured onboarding and long-term skill support reduces long-term adoption. Not all users will be "digital natives," and the system must accommodate a wide range of technical fluency levels. This is particularly relevant in clinical domains, where many users, especially senior medical professionals, may have limited exposure to emerging technologies or non-traditional interaction models. Unlike software developers or engineers, clinicians typically do not interact with 3D interfaces, virtual environments, or head-mounted displays in their day-to-day work. As a result, introducing AR tools into such environments requires careful preparation, onboarding, and gradual exposure to new interaction models. From a design perspective, this means anticipating resistance or uncertainty and embedding progressive learning scaffolds directly into the system. The goal is to reduce friction between the user's mental model, shaped largely by 2D interfaces, and the interaction model offered by AR, which often requires spatial reasoning, gesture memory, and a stronger sense of embodied interaction. Designing for graceful degradation, where the system can still provide value with minimal interaction, is one strategy to support less tech-savvy users without overwhelming them.

From an HCI standpoint, fostering learnability, the extent to which users can improve their performance with experience, is essential to long-term adoption [193]. Interfaces should become more efficient with repeated use, while retaining fallback guidance for occasional or returning users.

Ultimately, recognizing the learning curve as a core design concern, rather than an afterthought, enables smoother integration of AR systems into the medical domain. It also ensures that the technology empowers clinical users rather than placing additional cognitive or procedural burdens on them.

8.2. Clinical Environment and Workflow Integration

While usability and human-centered interaction are foundational to AR system design, they are insufficient on their own if the system cannot be meaningfully integrated into the dynamic and constrained environments of clinical practice.

The success of any augmented reality solution in the medical domain depends heavily on how well it fits into existing clinical workflows, tool ecologies, spatial constraints, and collaborative structures. Clinical environments are not static or generic, they are highly context-sensitive, time-pressured, and shaped by organizational norms, interdependencies, and physical limitations.

Designing for such environments requires attention to situated practice, workflow compatibility, and in-situ evaluation.

8.2.1. Situated Use and Context-Aware Design

A central principle in HCI is that user interaction is shaped not just by interface design but by situated action, the real-world context in which the action takes place. Suchman's foundational work argues that behavior emerges dynamically in context, rather than following predefined plans [244]. This insight is particularly salient in clinical environments, where the same treatment procedure or operation may be performed differently depending on patient anatomy, team composition, or unexpected incident developments.

AR systems must be designed not only for users but for the situations in which those users operate. Context-aware design in AR entails detecting and adapting to variables such as the phase of the procedure, the role of the user (e.g., lead surgeon vs. assistant), the available space, and concurrent device usage.

8.2.2. Workflow Compatibility and Seamless Integration

Clinical workflows are intricate and tightly choreographed. A single disruption, whether from misaligned technology, interface delays, or unclear output, can compromise safety or efficiency. Therefore, AR systems must be designed to augment clinical practice without interrupting it.

The SEIPS model (Systems Engineering Initiative for Patient Safety) developed by Carayon et al. [38] emphasizes the integration of tools, tasks, people, and organizational structures in healthcare environments. This model provides a useful lens for understanding how AR tools must be evaluated not in isolation, but as part of the larger sociotechnical system.

For example, an AR interface used in the operating room should integrate with existing imaging systems, avoid interfering with sterile fields, and respect the temporal pacing of procedures. Features such as hands-free interaction, voice control, and gesture-based navigation can facilitate this integration, but only if they are robust to environmental noise, surgical gloves, and movement constraints.

8.2.3. Spatial and Physical Constraints

Clinical settings are often characterized by spatial limitations, crowded team dynamics, and competing technologies. Unlike desktop computing environments, where the interaction space is controlled and predictable, clinical AR

applications must contend with shared physical spaces that are already saturated with equipment and personnel.

Isella et al. [117] demonstrate how spatial dynamics in hospital environments, captured through wearable sensors, expose dense, fast-paced interaction patterns among clinicians. These findings highlight how unpredictable spatial flows and human proximity patterns can complicate the deployment of sensor- or gesture-based AR systems, especially in multi-user environments like operating rooms or ICUs.

In the design of AR systems, this means avoiding obstructive visual projections and ensuring reliable performance across varying lighting conditions, user positions, and viewing angles, all while maintaining compatibility with sterile protocols. Furthermore, interaction modalities, such as gesture, gaze, or voice input, must be carefully adapted to the physical realities of clinical environments, including the use of gloved hands, ambient noise, and the need for unobtrusive operation within shared, high-traffic spaces.

8.2.4. Real-World Evaluation and Iterative Refinement

Effective integration cannot be validated solely in labs or simulated environments. It requires in-situ evaluation, testing and refining the AR system in real-world clinical workflows, with all their unpredictability and constraints.

The effective integration of AR systems into healthcare workflows requires more than controlled lab testing or high-fidelity simulations. While such settings are useful in early-stage development, they rarely capture the full ecological complexity and unpredictability of real-world clinical practice. In HCI research, this limitation is well recognized: systems often behave differently when removed from the tightly controlled lab and introduced into context-rich environments where human routines, technical infrastructure, and physical constraints interact in unpredictable ways [36, 66].

In-situ evaluation, the process of testing systems within the actual environments in which they will be used, offers a critical methodology for uncovering issues that remain invisible in abstracted settings. Such evaluations provide insight into situated action and emergent practices, how people adapt their use of a system in real time and how the system, in turn, shapes the flow of work [244]. For AR systems, in-situ testing is particularly essential due to their embodied interaction models, reliance on spatial fidelity, and sensitivity to real-time environmental input.

Our study with the ARAS system, an AR-based surgical navigation tool, exemplifies the value of in-situ, iterative testing. Early lab-based feedback was positive, but only through multiple rounds of field testing with surgical teams in real operating environments did subtle yet critical design issues emerge. Iterative refinement based on these findings allowed the ARAS system to gradually evolve into a more context-sensitive, workflow-aligned tool. This process reflects well-established principles in HCI, particularly those of participatory design and human-in-the-loop evaluation, which stress the importance of continuous feedback loops between system design and user practice [96].

Furthermore, in-situ evaluation fosters user trust and system credibility, two key factors in the adoption of novel technology in high-risk domains like healthcare. As noted by Blandford et al. [26], designed systems must be tested not only for usability, but also for their capacity to integrate seamlessly into existing work systems without compromising safety, efficiency, or professional autonomy.

In sum, deploying AR systems in healthcare settings requires more than technical excellence; it demands a deep commitment to contextual adaptation, empirical validation under real-world conditions, and responsiveness to the evolving needs and feedback of users. In-situ evaluation is not an optional step in this process, it is an essential practice for ensuring that AR technologies move beyond conceptual promise to practical, trustworthy, and impactful clinical tools.

8.3. Interdisciplinary Collaboration and the Imperative of Co-Creation

Developing and deploying AR systems in healthcare is not merely a technical challenge, it is a deeply interdisciplinary endeavor. It demands close, sustained collaboration across clinical practitioners (e.g., surgeons, nurses, radiologists), system designers, HCI researchers, regulatory experts, and often ethicists or patient advocates. Each group brings a unique perspective on what constitutes value, risk, and success.

This collaboration is not only desirable, it is essential. In clinical practice, the ultimate responsibility for patient safety and outcomes rests with the healthcare professionals. If these users are expected to rely on AR systems in time-critical, high-consequence environments, they must also be empowered to evaluate, question, and understand those systems. However, such evaluation is only possible when users trust the underlying design and development process, a trust that cannot be achieved through documentation or training alone.

As our own development work has also shown, trust is significantly enhanced when clinical users are involved as co-creators, not merely as testers or end-users. When clinicians actively contribute to shaping system behavior, visualizations, and interaction models, they not only improve the system design and quality but also develop a deeper sense of ownership and confidence in the technology.

This dynamic was evident throughout our study with ARAS. As surgeons gained familiarity with the AR system and experienced its positive impact on surgical precision and efficiency, their trust and willingness to use and contribute to improving the technology grew. This observed increase in user acceptance correlated with surgeons consistently reporting ARAS's usefulness during the postoperative interviews. This observation also aligns with davis1989perceived's work, which shows that perceived usefulness strongly correlates with user behavior. This chain of influence reflects a well-established pathway in HCI and technology acceptance research, where participatory in-

volvement supports both system credibility and long-term integration into practice.

In many cases, however, the collaboration between technical developers and clinical staff is hindered by disciplinary silos, terminology gaps, or mismatched incentives. Addressing these challenges requires creating shared spaces, conceptually and practically, where mutual learning and joint decision-making can occur. Methods from participatory design, action research, and reflective practice in HCI provide effective frameworks for enabling this kind of interdisciplinary engagement [237].

Moreover, interdisciplinary research contributes to research legitimacy itself. Clinical users are more likely to commit their limited time and professional credibility to validating a system if they see their concerns reflected in its assumptions and design choices. In contrast, systems developed without clinical input often face invisible resistance, where users opt out of real-world testing or undermine deployment through passive non-engagement.

In short, the co-creation of AR systems is not just a design strategy, it is a mechanism for building trust, ensuring responsibility, and enabling robust evaluation in the high-stakes context of clinical care.

8.4. Ethical Considerations

Systems designed for use in healthcare inherently carry risks, particularly when they influence clinical decisions, expose sensitive patient data, or operate semi-autonomously. For AR to gain legitimacy and trust in the clinical domain, its design and implementation must be guided by principles of responsible innovation, value-sensitive design, and regulatory alignment.

8.4.1. Embedding Ethical Values Through Design

AR systems mediate perception, decision-making, and behavior. As such, they are not ethically neutral tools, but active agents in shaping clinical practice. The HCI concept of Value Sensitive Design (VSD) provides a structured approach to embedding human values, such as privacy, accountability, and autonomy, into the design process from the outset [89]. Unlike approaches that treat ethics as an afterthought, VSD emphasizes proactive ethical deliberation during requirements gathering, prototyping, and evaluation.

For example, in surgical AR systems, decisions must be made about which anatomical structures are visualized, when alerts are triggered, and who controls the display. These choices can have ethical implications, for instance, by reinforcing hierarchical control, obscuring critical information, or unintentionally biasing the user toward a particular decision path.

Involving diverse stakeholders (e.g., surgeons, nurses, ethicists, IT staff) in participatory design activities is key to ensuring that such systems reflect the moral and professional values of their intended users.

8.4.2. Algorithmic Transparency and the Risk of Over-Reliance

As AR systems increasingly incorporate intelligent and data-driven components, including automated annotation, predictive modeling, and computer vision, the issue of algorithmic opacity becomes a central ethical and practical concern. Systems that interpret and display complex patient data through AR interfaces can inadvertently shape clinical decision-making, sometimes without clearly communicating their reasoning or limitations. Mittelstadt et al. [183] highlight how such systems may obscure biases, amplify errors, and create accountability gaps when their outputs are not transparent or contestable.

In clinical AR, these risks are compounded by the perceptual authority of visual overlays. When a system highlights a vessel, or prioritizes one anatomical region over another, clinicians may feel compelled to follow its guidance, even in the absence of clear justifications. This phenomenon, known as automation bias, can undermine critical thinking and lead to over-reliance, particularly in users with lesser clinical and surgical expertise.

When embedded within AR systems, LLMs have the potential to enhance interpretability and user trust by delivering natural language explanations, generating contextual summaries, or serving as an intuitive bidirectional communication interface. As demonstrated in our work, LLMs can simultaneously simplify system design, by consolidating multiple input/output pathways, and support more natural, human-centered interaction. For example, a clinician might ask, "Why is this region highlighted?", and receive a concise explanation grounded in anatomical context or statistical inference. This functionality supports real-time sensemaking and aligns closely with emerging research in explainable AI (XAI) and interactive transparency, which emphasizes the need for systems that foster human oversight, comprehension, and agency [7, 64].

However, LLMs also carry risks. Their outputs may appear fluent and confident while being factually incorrect or contextually misaligned with clinical realities. The ambiguity in source attribution, lack of standardization in explanation quality, and potential for hallucinations present significant barriers to safe integration in clinical workflows. Without proper safeguards, LLM-generated feedback may reinforce over-reliance rather than mitigate it, particularly when embedded in visually authoritative AR environments.

To address these challenges, beyond implementing strong safeguards, they must include transparency scaffolding: visual or verbal indicators of uncertainty, options for verifying model outputs, and access to underlying data or rationale. Effective automation must not only be technically accurate but also communicative about its confidence levels, assumptions, and boundaries. Embedding these features within AR interfaces, without overwhelming or distracting the user, requires careful interaction design and continuous evaluation in situ.

8.4.3. Patient-Centered Ethical Considerations

While much of the discussion around AR system design in healthcare focuses on technical feasibility and clinical usability, the primary ethical responsibility

lies with safeguarding the rights, well-being, and dignity of the patient. Any intervention, technological or otherwise, that influences diagnosis, treatment, or clinical decision-making must be designed with the patient as the ultimate stakeholder. This includes ensuring patient safety, data privacy, consent, and the avoidance of harm, in both direct and indirect forms.

Safety is a fundamental concern. AR systems that misrepresent information, delay decision-making, or introduce visual or cognitive distractions can compromise clinical outcomes. Because these systems are often integrated into time-critical and high-risk workflows, even small usability failures can have serious consequences.

Privacy is another cornerstone of patient-centered ethics, particularly given the data-intensive nature of AR systems. These platforms may access or visualize sensitive health information, imaging scans, physiological signals, or real-time biometric data, raising concerns about data exposure, unauthorized access, and secondary use. Developers must implement privacy-by-design principles, including strong encryption, role-based access control, and transparency in data handling practices [89, 183].

Moreover, informed consent becomes increasingly complex in AR-enhanced settings. Patients may not fully understand what data is being collected, how it is being visualized, or who has access to it, especially in contexts involving telepresence, mixed-reality training, or AI-generated overlays. Ensuring comprehensible and context-specific consent procedures is crucial, as is offering patients meaningful choices about how their data is used and shared.

Finally, ethical design must also account for patient dignity and the human experience of care. Technologies that reduce patients to data points, diminish interpersonal interaction, or overly medicalize routine procedures may inadvertently dehumanize the care experience. AR designers should therefore strive to create systems that are not only functional and efficient but also respectful, empathetic, and transparent in how they intersect with the patient's body and narrative.

8.5. Generalizability and Domain-Specific Constraints

A recurring challenge in the design of AR systems for healthcare is the question of generalizability: to what extent can findings, methods, or interaction patterns developed in one clinical context be transferred to another, or even to different domains entirely? From a HCI perspective, this challenge reveals the tension between universal design principles and the situated nature of use.

Traditional HCI has often sought to develop reusable design patterns and heuristics that apply across settings, promoting scalability, efficiency, and consistency. However, as Suchman's theory of situated action emphasizes, user behavior and system use are heavily shaped by context, including physical environments, professional roles, organizational norms, and temporal pressures [244]. In healthcare, these contextual variables are especially pronounced: different clinical specialties, institutional cultures, and even surgical teams may exhibit divergent practices, mental models, and risk tolerances.

In our work with different medical specialties, it became evident that interaction modalities, information needs, and even trust dynamics varied significantly, not only across institutions, but among individuals with different clinical roles and levels of experience. This reflects a broader pattern noted in HCI literature: while generalizable frameworks (e.g., participatory design, value-sensitive design) can guide development, interface-level generalizations are often inappropriate or misleading in safety-critical domains [26, 51].

Moreover, the medical domain imposes unique constraints that further limit generalizability. Unlike consumer or enterprise software, clinical systems must align with regulatory standards, integrate into highly structured workflows, and accommodate responsibilities that carry legal and ethical weight. Design decisions that succeed in one surgical workflow may prove infeasible or unsafe in another due to differences in available technology, staff roles, or procedural pacing. In this sense, seeking a "universal" AR solution for healthcare is not only impractical, it may be actively counterproductive.

Instead, researchers and developers must approach clinical AR system design with a focus on contextual fit, iterative adaptation, and local validation. HCI practices such as in-situ evaluation, design ethnography, and reflexive participatory methods offer more appropriate paths forward than one-size-fits-all toolkits or interface templates.

In conclusion, while some foundational design insights may transcend domains, the development of clinical AR systems must embrace situatedness, specificity, and responsiveness as core design values. Attempting to generalize without grounding solutions in real-world clinical variation risks overlooking the very complexities that define healthcare practice.

Chapter 9

Conclusion and Future Directions

9.1. Summary of Contributions

This dissertation investigates how augmented reality can be harnessed to improve human performance in tasks that demand substantial cognitive effort, precise motor control, and rapid decision-making under pressure. Critical medical procedures, in which these challenges are inherently present, serve as the primary context for this investigation. Through an interdisciplinary approach that integrates computer science and medical practice, the work develops methods designed to support clinicians in high-demand environments.

Rather than approaching AR in healthcare from a single disciplinary perspective, the thesis adopts a mixed methods framework that combines engineering design, human-computer interaction, real-world clinical evaluation, and outcome-based analysis. As a result, it provides a cohesive and practically oriented contribution that addresses both the technical foundations and the human and clinical contexts necessary for the meaningful adoption of AR technologies in healthcare.

The work begins by establishing foundational knowledge for the medical domain, investigating how hardware and software-related factors influence human performance and usability in AR systems. This contribution integrates literature review, empirical evaluation, and user-centered studies to identify key constraints that inform design and deployment decisions (Chapter 2 and Chapter 3). It further advances system engineering and development through two implemented applications: one to support the acquisition of procedural and motor skills (Chapter 4), and another for preoperative and intraoperative cognitive support during pancreatic surgery (Chapter 5). Together, these systems illustrate scalable engineering and design approaches to enhance user performance using AR in cognitively challenging and time-pressured domains across an extended spectrum from training and skill acquisition to real-time cognitive support within surgical contexts. By exploring the design space for support under cognitively demanding environments such as intraoperative scenarios, the thesis also proposes an innovative use of large language models to enable context-aware system design that reduces cognitive load, accelerates

task completion, and simplifies system development (Chapter 7).

In addition, the thesis provides a rigorous evaluation of these methods in ecologically valid environments. By conducting clinical trials, it demonstrates the downstream real-world impact of augmented reality technologies on both surgical performance and patient outcomes (Chapter 6). Finally, in its endeavor to expand knowledge in human-computer interaction for the medical domain, the thesis synthesizes insights from interdisciplinary collaboration, user feedback, and HCI theory to propose a holistic framework for design in safety-critical contexts, emphasizing co-creation, workflow integration, human-centered design, and ethical considerations (Chapter 8).

Taken together, these contributions form an integrated framework that advances augmented reality in medicine from conceptual innovation to practical, clinically deployable solutions. By uniting system engineering, adaptive intelligence, human-centered design, and empirical validation, this work provides a replicable model for the future development of augmented reality technologies in healthcare, demonstrating how interdisciplinary approaches are essential to achieving both technical excellence and clinical relevance.

9.2. Future Works

As AR technologies mature and begin to find applications in high-stakes medical environments, several critical directions emerge for future research. One of the most pressing gaps in the development of AR systems for medical applications is the absence of domain-specific benchmarks and standardized evaluation protocols. The pursuit of perfection of technical performance when it comes to the medical field, in the absence of context-aware benchmarks, risks obscuring the actual value AR technologies bring to practice. Current evaluations tend to emphasize raw technical metrics such as spatial accuracy, latency, and visual fidelity, often without sufficient consideration of the clinical context in which the system will be deployed.

While technical excellence is undoubtedly important, there is a common misconception that perfection, typically defined in terms such as millimetric accuracy for virtual object registration or flawless system stability, is universally required across all medical domains. This assumption overlooks the fact that what constitutes “perfection” or even “usability” can vary significantly between and even within clinical fields.

A system that slightly compromises on precision but enhances clinician situational awareness, reduces procedure time, or supports better training outcomes may be far more impactful than one that achieves near-perfect tracking in a lab but fails in real-world adoption. Without domain-specific standards, such trade-offs are difficult to quantify or justify, leading to both underreporting of meaningful benefits and overemphasis on technical refinements that may not translate to clinical utility.

Future research should therefore focus on establishing domain-specific benchmarks that align performance evaluation with the specific goals, risks, and workflows of different medical contexts. Such benchmarks should be devel-

oped in collaboration with clinicians and domain experts to ensure clinical relevance, and they should include not only system-centric metrics, such as latency, tracking stability, or field of view, but also human-centric measures such as task efficiency, error reduction, cognitive load, and decision confidence. Equally important is the need to standardize measurement approaches. At present, the variety of tools, methodologies, and protocols used to assess AR systems leads to fragmented results and complicates comparison. Agreement on how performance is measured will be essential for building a coherent evidence base.

Moving forward, the field must recognize that effectiveness in medical AR is not universally defined by extreme precision. At the same time, impact should be assessed in relation to the specific clinical function the system serves. Establishing robust, domain-appropriate benchmarks and standardizing measurement approaches are critical next steps toward ensuring that AR technologies are not only technically impressive but also clinically meaningful, scalable, and trusted in practice.

Complementing the need for standardized evaluation, another promising direction for future research is the development of personalized and intelligent AR systems that adapt dynamically to individual users and clinical contexts. Most existing medical systems are designed with a one-size-fits-all approach, offering little adaptation to the user's experience level, task complexity, or situational demands. As a result, the burden typically falls on the user to learn and adapt to the system, rather than the system accommodating the user. However, clinical environments are inherently variable, and users differ widely in cognitive load tolerance, procedural familiarity, and interaction preferences. Future AR systems should be capable of tailoring content delivery and interaction modalities based on user profiles, such as expertise level (e.g., trainee vs. senior surgeon), physiological states (e.g., stress or fatigue indicators), or past performance. The integration of intelligent agents, including large language models, further opens the possibility for context-aware guidance, decision support, and real-time adaptation. For instance, an intelligent AR system could offer more detailed step-by-step instructions to a novice, while providing high-level cues or alerts to an expert. It could also detect procedural deviations or user hesitation and respond proactively with timely prompts or additional resources. Developing such personalized and adaptive systems will require interdisciplinary research across human-computer interaction, machine learning, and clinical informatics. Key challenges include maintaining user trust, avoiding over-reliance on automation, and ensuring that the system adapts in ways that enhance rather than distract from clinical performance. Nevertheless, personalization and intelligent responsiveness represent a critical frontier in making AR not just an overlay technology, but a truly integrated and supportive part of clinical decision-making and performance.

Altogether, personalization and intelligent responsiveness represent a vital step forward in making AR not just a visual augmentation tool, but a deeply integrated and supportive element of clinical decision-making and performance. When paired with robust benchmarking and evaluation standards,

these future directions promise to move the field beyond technical demonstration toward truly impactful, scalable, and safe clinical deployment.

Bibliography

[1] Paola Adamo, Gianluca Longhi, Federico Temporiti, Giorgia Marino, Emilia Scalona, Maddalena Fabbri-Destro, Pietro Avanzini, and Roberto Gatti. "Effects of Action Observation Plus Motor Imagery Administered by Immersive Virtual Reality on Hand Dexterity in Healthy Subjects". In: *Bioengineering* 11.4 (2024), p. 398.

[2] Pedro Aguilar-Salinas, Salvador F Gutierrez-Aguirre, Mauricio J Avila, and Peter Nakaji. "Current status of augmented reality in cerebrovascular surgery: A systematic review". In: *Neurosurgical Review* 45.3 (2022), pp. 1951–1964.

[3] F Ahmad, W Ahmad, and J Xiong. "AR and MR in Dentistry: Developments, Applications, and Prospects". In: *IEEE Transactions on Medical Imaging* (2024). URL: `https://ieeexplore.ieee.org/document/10759859`.

[4] E Ajitha and S Umamaheswari. "Real-Time Health Assessment Using AR for Triage". In: *IEEE Conference Proceedings* (2025). URL: `https://ieeexplore.ieee.org/abstract/document/10933184/`.

[5] Mareen Allgaier, Vuthea Chheang, Patrick Saalfeld, Vikram Apilla, Tobias Huber, Florentine Huettl, Belal Neyazi, I Erol Sandalcioglu, Christian Hansen, Bernhard Preim, et al. "A comparison of input devices for precise interaction tasks in VR-based surgical planning and training". In: *Computers in Biology and Medicine* 145 (2022), p. 105429.

[6] Ibraheem Alrishan Alzouebi, Sanad Saad, Tom Farmer, and Sophie Green. "Is the use of augmented reality-assisted surgery beneficial in urological education? A systematic review". In: *Current Urology* 15.3 (2021), p. 148.

[7] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. "Guidelines for human-AI interaction". In: *Proceedings of the 2019 chi conference on human factors in computing systems*. 2019, pp. 1–13.

[8] Misuk An, Youngmee Kim, and Won-Kyung Cho. "Effect of smart devices on the quality of CPR training: A systematic review". In: *Resuscitation* 144 (2019), pp. 145–156.

[9] PE Antoniou, G Arfaras, and N et al. Pandria. "Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: cohort study". In: *JMIR Serious Games* (2020). URL: https://www.researchprotocols.org/2020/3/e17823/.

[10] Apple. *Apple Vision Pro*. 2025. URL: https://www.apple.com/de/shop/buy-vision/apple-vision-pro.

[11] Apple Inc. *Apple Vision Pro Technical Specifications*. Website. Accessed: 2025-03-15. 2025. URL: https://www.apple.com/apple-vision-pro/specs/.

[12] Arduino. *Arduino Pro mini*. 2018. URL: www.arduino.cc.

[13] World Medical Association. *WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants*. https://www.wma.net/policies-post/wma-declaration-of-helsinki/. Adopted at the 75th WMA General Assembly, Helsinki, Finland. Oct. 2024.

[14] ATLASti. *ATLAS.ti: The Qualitative Data Analysis and Research Software*. Accessed: 2024-09-06. 2023. URL: https://atlasti.com/.

[15] Yahia Baashar, Gamal Alkawsi, Wan Nooraishya Wan Ahmad, Hitham Alhussian, Ayed Alwadain, Luiz Fernando Capretz, Areej Babiker, and Adnan Alghail. "Effectiveness of using augmented reality for training in the medical professions: meta-analysis". In: *JMIR serious games* 10.3 (2022), e32715.

[16] S Balian, SK McGovern, BS Abella, and AL Blewer. "Feasibility of an Augmented Reality Cardiopulmonary Resuscitation Training System". In: *Heliyon* (2019). URL: https://www.cell.com/heliyon/fulltext/S2405-8440(19)35865-7.

[17] Steve Balian, Shaun K McGovern, Benjamin S Abella, Audrey L Blewer, and Marion Leary. "Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers". In: *Heliyon* 5.8 (2019), e02205.

[18] Aaron Bangor, Philip Kortum, and James Miller. "Determining what individual SUS scores mean: Adding an adjective rating scale". In: *Journal of usability studies* 4.3 (2009), pp. 114–123.

[19] EZ Barsom, M Graafland, and MP Schijven. "Systematic review on the effectiveness of augmented reality applications in medical training". In: *Surgical Endoscopy* (2016). URL: https://link.springer.com/article/10.1007/s00464-016-4800-6.

[20] Robert A Berg, Karl B Kern, Ronald W Hilwig, Marc D Berg, Arthur B Sanders, Charles W Otto, and Gordon A Ewy. "Assisted ventilation does not improve outcome in a porcine model of single-rescuer bystander cardiopulmonary resuscitation". In: *Circulation* 95.6 (1997), pp. 1635–1641.

[21] H Bhatti, MMA Bhatti, and M et al. Ali. "Augmented and Virtual Reality in Radiology: A Comparative Analysis of Clinical Applications and Outcomes". In: *Journal of Medical Health Sciences Research* (2025). URL: https://jmhsr.com/index.php/jmhsr/article/view/159.

[22] EG Bignami, M Russo, F Semeraro, and V Bellini. "Technoethics in Clinical AR Education". In: *Springer Critical Care and Anesthesia* (2025). URL: https://link.springer.com/content/pdf/10.1186/s44158-025-00233-2.pdf.

[23] Emrullah Birgin, Sebastian Hempel, Alina Reeg, Florian Oehme, Annika Schnizer, Johann S Rink, Matthias F Froelich, Svetlana Hetjens, Verena Plodeck, Heiner Nebelung, et al. "Development and validation of a model for postpancreatectomy hemorrhage risk". In: *JAMA Network Open* 6.12 (2023), e2346113–e2346113.

[24] M Birlo, PJE Edwards, M Clarkson, and D Stoyanov. "Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review". In: *Medical Image Analysis* (2022). URL: https://www.sciencedirect.com/science/article/pii/%20S1361841522000147.

[25] Manuel Birlo, PJ Eddie Edwards, Matthew Clarkson, and Danail Stoyanov. "Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review". In: *Medical Image Analysis* 77 (2022), p. 102361.

[26] Ann Blandford, Dominic Furniss, and Stephann Makri. *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers, 2016.

[27] Jakub Blokša. "Design guidelines for user interface for augmented reality". In: *Masaryk University* (2017).

[28] Simone Borsci, Peter Buckle, and Simon Walne. "Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology?" In: *Applied ergonomics* 84 (2020), p. 103007.

[29] Allan Braslow, Robert T Brennan, Mary M Newman, Nicholas G Bircher, Anne M Batcheller, and William Kaye. "CPR training without an instructor: development and evaluation of a video self-instructional system for effective performance of cardiopulmonary resuscitation". In: *Resuscitation* 34.3 (1997), pp. 207–220.

[30] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[31] Todd B Brown, Johnny A Dias, Devashish Saini, Reesha C Shah, Stacey S Cofield, Thomas E Terndrup, Richard A Kaslow, and John W Waterbor. "Relationship between knowledge of cardiopulmonary resuscitation guidelines and performance". In: *Resuscitation* 69.2 (2006), pp. 253–261.

[32] Raphael Romano Bruno, Georg Wolff, Bernhard Wernly, Maryna Masyuk, Kerstin Piayda, Susannah Leaver, Ralf Erkens, Daniel Oehler, Shazia Afzal, Houtan Heidari, et al. "Virtual and augmented reality in critical care medicine: the patient's, clinician's, and researcher's perspective". In: *Critical Care* 26.1 (2022), p. 326.

[33] Lorrie A Buddenberg and Chris Davis. "Test–retest reliability of the Purdue Pegboard Test". In: *The American Journal of Occupational Therapy* 54.5 (2000), pp. 555–558.

[34] T Bui, MA Ruiz-Cardozo, and MR et al. Kann. "Virtual, augmented, and mixed reality applications for surgical rehearsal, operative execution, and patient education in spine surgery: a scoping review". In: *Medicina* (2024). URL: https://www.mdpi.com/1648-9144/60/2/332.

[35] A Burlacu, C Brinza, and NN Horia. "How the Metaverse Is Shaping the Future of Healthcare Communication". In: *Cureus* (2025). URL: https://www.cureus.com/articles/351003.

[36] Bill Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan kaufmann, 2010.

[37] Kelly Caine. "Local standards for sample size at CHI". In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 981–992.

[38] PASH Carayon, A Schoofs Hundt, Ben-Tzion Karsh, Ayse P Gurses, Carla J Alvarado, Michael Smith, and P Flatley Brennan. "Work system design for patient safety: the SEIPS model". In: *BMJ Quality & Safety* 15.suppl 1 (2006), pp. i50–i58.

[39] João Cartucho, David Shapira, Hutan Ashrafian, and Stamatia Giannarou. "Multimodal mixed reality visualisation for intraoperative surgical guidance". In: *International journal of computer assisted radiology and surgery* 15 (2020), pp. 819–826.

[40] WA Chatha. "From Scalpel to Simulation: Reviewing the Future of Cadaveric Dissection in the Upcoming Era of Virtual and Augmented Reality and Artificial Intelligence". In: *Cureus* (2024). URL: https://www.cureus.com/articles/302914.

[41] E Checcucci, S Puliatti, G Volpi, and G Bignante. "Harnessing Advanced Imaging Technologies in Robotic Urological Surgery Requires Specialized and Structured Training". In: *European Urology Open Science* (2025). URL: https://www.sciencedirect.com/science/article/pii/S266616832500120X.

[42] Jia-Hao Chen, Li-Yong Zhu, Zhi-Wei Cai, Xiao Hu, Abousalam Abdoulkader Ahmed, Jie-Qiong Ge, Xiao-Yan Tang, Chun-Jing Li, Yun-Long Pu, and Chong-Yi Jiang. "TRIANGLE operation, combined with adequate adjuvant chemotherapy, can improve the prognosis of pancreatic head cancer: a retrospective study". In: *World Journal of Gastrointestinal Oncology* 16.5 (2024), p. 1773.

[43] Hang Cheng, Jeffrey W Clymer, Brian Po-Han Chen, Behnam Sadeghirad, Nicole C Ferko, Chris G Cameron, and Piet Hinoul. "Prolonged operative duration is associated with complications: a systematic review and meta-analysis". In: *Journal of Surgical Research* 229 (2018), pp. 134–144.

[44] N Cheng, H Wang, Y Ma, J Chen, Y Zhang, and RS Sherratt. "Augmented reality in healthcare applications: A review". In: *Artificial Intelligence in Medicine* (2020). URL: https://www.sciencedirect.com/science/article/pii/S0933365719304001.

[45] Ki-Hyun Cho, Francis A Papay, Jeff Yanof, Karl West, Bahar Bassiri Gharb, Antonio Rampazzo, Brian Gastman, and Graham S Schwarz. "Mixed reality and 3D printed models for planning and execution of face transplantation". In: *Annals of Surgery* 274.6 (2021), e1238–e1246.

[46] Mathias Ciliberto, Vitor Fortes Rey, Alberto Calatroni, Paul Lukowicz, and Daniel Roggen. "Opportunity++: A multimodal dataset for video-and wearable, object and ambient sensors-based human activity recognition". In: *Frontiers in Computer Science* 3 (2021), p. 792065.

[47] Enylton Machado Coelho, Blair MacIntyre, and Simon J Julier. "Supporting interaction in augmented reality in the presence of uncertain spatial knowledge". In: *Proceedings of the 18th annual ACM symposium on User interface software and technology*. 2005, pp. 111–114.

[48] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. routledge, 2013.

[49] Rachel Cooper, Sabine Junginger, and Thomas Lockwood. *The handbook of design management*. Bloomsbury Publishing, 2017.

[50] IBM Corp. *IBM SPSS Statistics for Windows (Version 22.0)*. Armonk, NY, IBM Corp, 2017. URL: www.ibm.com/products/spss-statistics.

[51] Andrew Crabtree, Mark Rouncefield, and Peter Tolmie. *Doing design ethnography*. Springer Science & Business Media, 2012.

[52] Fabrizio Cutolo, Benish Fida, Nadia Cattari, and Vincenzo Ferrari. "Software framework for customized augmented reality headsets in medicine". In: *IEEE Access* 8 (2019), pp. 706–720.

[53] Fabrizio Cutolo, Paolo Domenico Parchi, and Vincenzo Ferrari. "Video see through AR head-mounted display for medical procedures". In: *2014 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE. 2014, pp. 393–396.

[54] Xiang Da Dong, Daniel Moritz Felsenreich, Shekhar Gogna, Aram Rojas, Ethan Zhang, Michael Dong, Asad Azim, and Mahir Gachabayov. "Robotic pancreaticoduodenectomy provides better histopathological outcomes as compared to its open counterpart: a meta-analysis". In: *Scientific reports* 11.1 (2021), p. 3774.

[55] Shaveta Dargan, Shally Bansal, Munish Kumar, Ajay Mittal, and Krishan Kumar. "Augmented reality: A comprehensive review". In: *Archives of Computational Methods in Engineering* 30.2 (2023), pp. 1057–1080.

[56] Ryan D Datteri and Benoıt M Dawant. "Estimation and reduction of target registration error". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part III 15*. Springer. 2012, pp. 139–146.

[57] Philipp Dautzenberg, Holger A Volk, Nikolaus Huels, Lena Cieciora, Katharina Dohmen, Matthias Lüpke, Herman Seifert, and Oliver Harms. "The effect of steam sterilization on different 3D printable materials for surgical use in veterinary medicine". In: *BMC veterinary research* 17 (2021), pp. 1–6.

[58] F Demir and P Calyam. "A Next-Generation Augmented Reality Platform for Mass Casualty Incidents". In: *Journal of Usability Studies* (2017). URL: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Demir_August2017.pdf.

[59] Johanne Desrosiers, Réjean Hébert, Gina Bravo, and Elisabeth Dutil. "The Purdue Pegboard Test: normative data for people aged 60 and over". In: *Disability and rehabilitation* 17.5 (1995), pp. 217–224.

[60] A Dias, D Männle, T Balkenhol, Jürgen Hesser, N Rotter, L Huber, O Hoffmann, A Schell, B Kramer, A Lammert, et al. "Augmented Reality during Parotid Surgery: Real-Life Evaluation of Voice Control and User-Experience". In: *Laryngo-Rhino-Otologie* 100.S 02 (2021).

[61] RM Dickey, N Srikishen, and LI Lipshultz. "Augmented reality assisted surgery: a urologic training tool". In: *Asian Journal of Andrology* (2016). URL: https://journals.lww.com/ajandrology/fulltext/2016/18050/Augmented_reality_assisted_surgery__a_urologic.12.aspx.

[62] Marcello Donati, Francesco Basile, Gregor A Stavrou, and Karl J Oldhafer. "Navigation systems in liver surgery: the new challenge for surgical research". In: *Journal of Laparoendoscopic & Advanced Surgical Techniques* 23.4 (2013), pp. 372–375.

[63] Xin Luna Dong, Seungwhan Moon, Yifan Ethan Xu, Kshitiz Malik, and Zhou Yu. "Towards next-generation intelligent assistants leveraging llm techniques". In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, pp. 5792–5793.

[64] Finale Doshi-Velez and Been Kim. "Towards a rigorous science of interpretable machine learning". In: *arXiv preprint arXiv:1702.08608* (2017).

[65] MWG Doughty, NR Ghugre, and GA Wright. "Augmenting performance: A systematic review of optical see-through head-mounted displays in surgery". In: *Journal of Imaging* (2022). URL: https://www.mdpi.com/2313-433X/8/7/203.

[66] Paul Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2001.

[67] Anne-Lise Doyen and Michèle Carlier. "Measuring handedness: A validation study of Bishop's reaching card test". In: *Laterality: Asymmetries of Body, Brain and Cognition* 7.2 (2002), pp. 115–130.

[68] Sulaman Durrani, Chiduziem Onyedimma, Ryan Jarrah, Atiq Bhatti, Karim Rizwan Nathani, Archis R Bhandarkar, William Mualem, Abdul Karim Ghaith, Cameron Zamanian, Giorgos D Michalopoulos, et al. "The virtual vision of neurosurgery: how augmented reality and virtual reality are transforming the neurosurgical operating room". In: *World Neurosurgery* 168 (2022), pp. 190–201.

[69] Heinrich-Heine-Universität Düsseldorf. *G*Power: Statistical Power Analyses for Windows and Mac*. Accessed: 2025-04-09. 2025. URL: `https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower`.

[70] Faruk Lawal Ibrahim Dutsinma, Debajyoti Pal, Suree Funilkul, and Jonathan H Chan. "A systematic review of voice assistant usability: An ISO 9241–11 approach". In: *SN computer science* 3.4 (2022), p. 267.

[71] PJ " Eddie" Edwards, Manish Chand, Manuel Birlo, and Danail Stoyanov. "The challenge of augmented reality in surgery". In: *Digital Surgery* (2021), pp. 121–135.

[72] Jan Egger, Christina Gsaxner, Xiaojun Chen, Jiang Bian, Jens Kleesiek, and Behrus Puladi. "Apple Vision Pro for Healthcare:" The Ultimate Display"?–Entering the Wonderland of Precision Medicine". In: *arXiv preprint arXiv:2308.04313* (2023).

[73] Jan Egger, Christina Gsaxner, Gijs Luijten, Jianxu Chen, Xiaojun Chen, Jiang Bian, Jens Kleesiek, Behrus Puladi, et al. "Is the Apple Vision Pro the Ultimate Display? A First Perspective and Survey on Entering the Wonderland of Precision Medicine". In: *JMIR Serious Games* 12.1 (2024), e52785.

[74] Mica R Endsley. "Toward a theory of situation awareness in dynamic systems". In: *Human factors* 37.1 (1995), pp. 32–64.

[75] E Esmaeeli, A Soleimani, and MS Khorashadizadeh. "VR-Based Anatomy and Emergency Education Framework". In: *Scientific Reports* (2025). URL: `https://www.nature.com/articles/s41598-025-96074-8.pdf`.

[76] Ethicon. *PROLENE™ Suture Material*. 2025. URL: `https://www.jnjmedtech.com/de-DE/product/prolene-nahtmaterial`.

[77] Timothy Everson, Matthew Joordens, Helen Forbes, and Ben Horan. "Virtual reality and haptic cardiopulmonary resuscitation training approaches: A review". In: *IEEE Systems Journal* (2021).

[78] SM Faisal, M Ishrat, and W Khan. "Digital Twins in Healthcare Emergency Training". In: *Taylor Francis* (2025). URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003510338-4/digital-twins-healthcare.

[79] Cristina R Ferrone, Giovanni Marchegiani, Theodore S Hong, David P Ryan, Vikram Deshpande, Erin I McDonnell, Francesco Sabbatino, Daniela Dias Santos, Jill N Allen, Lawrence S Blaszkowsky, et al. "Radiological and surgical implications of neoadjuvant treatment with FOLFIRINOX for locally advanced and borderline resectable pancreatic cancer". In: *Annals of surgery* 261.1 (2015), pp. 12–17.

[80] B Fida, F Cutolo, G di Franco, M Ferrari, and V Ferrari. "Augmented reality in open surgery". In: *Updates in Surgery* (2018). URL: https://link.springer.com/article/10.1007/s13304-018-0567-8.

[81] Benish Fida, Fabrizio Cutolo, Gregorio di Franco, Mauro Ferrari, and Vincenzo Ferrari. "Augmented reality in open surgery". In: *Updates in surgery* 70.3 (2018), pp. 389–400.

[82] Marc Fischer, Christoph Leuze, Stephanie Perkins, Jarrett Rosenberg, Bruce Daniel, and Alejandro Martin-Gomez. "Evaluation of different visualization techniques for perception-based alignment in medical AR". In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2020, pp. 45–50.

[83] J Michael Fitzpatrick. "Fiducial registration error and target registration error are uncorrelated". In: *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*. Vol. 7261. SPIE. 2009, pp. 21–32.

[84] J Michael Fitzpatrick, Jay B West, and Calvin R Maurer. "Predicting error in rigid-body point-based registration". In: *IEEE transactions on medical imaging* 17.5 (1998), pp. 694–702.

[85] A Follmann, M Ohligs, and N Hochhausen. "Smart Glasses in Mass Casualty Triage Simulation". In: *JMIR Medical Informatics* (2019). URL: https://www.jmir.org/2019/1/e11939/.

[86] A Follmann, A Ruhl, M Gösch, and M Felzen. "Augmented reality for guideline presentation in medicine: Randomized crossover simulation trial". In: *JMIR mHealth and uHealth* (2021). URL: https://mhealth.jmir.org/2021/10/e17472.

[87] Andreas Follmann, Alexander Ruhl, Michael Gösch, Marc Felzen, Rolf Rossaint, and Michael Czaplik. "Augmented reality for guideline presentation in medicine: Randomized crossover simulation trial for technically assisted decision-making". In: *JMIR mHealth and uHealth* 9.10 (2021), e17472.

[88] Thomas Franke, Christiane Attig, and Daniel Wessel. "A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale". In: *International Journal of Human–Computer Interaction* 35.6 (2019), pp. 456–467.

[89] Batya Friedman, Peter H Kahn, Alan Borning, and Alina Huldtgren. "Value sensitive design and information systems". In: *Early engagement and new technologies: Opening up the laboratory* (2013), pp. 55–95.

[90] M Fritz, N Navab, and K Gavaghan. "In-situ visualization in image-guided surgery". In: *Computerized Medical Imaging and Graphics* (2014). URL: https://www.sciencedirect.com/science/article/pii/S0895611113001257.

[91] Boris A Galitsky. "Truth-o-meter: Collaborating with llm in fighting its hallucinations". In: (2023).

[92] G Ganesan and S Poonkuntran. "Semantic Segmentation Based on Geometric Calibration Using AI and AR in Health Care". In: *Contemporary Mathematics* (2024). URL: https://ojs.wiserpub.com/index.php/CM/article/view/4218.

[93] Yahya Ghazwani and Shamus Smith. "Interaction in augmented reality: Challenges to enhance user experience". In: *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations*. 2020, pp. 39–44.

[94] Nicolas Golse, Antoine Petit, Maıté Lewin, Eric Vibert, and Stéphane Cotin. "Augmented reality during open liver surgery using a markerless non-rigid registration system". In: *Journal of Gastrointestinal Surgery* 25.3 (2021), pp. 662–671.

[95] J Gowthamy, A Senthilselvi, Aniket Kumar, S Aakash, and Gandikota Sreedhar. "Enhanced AI Voice Assistance using Machine Learning and NLP". In: *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)*. Vol. 1. IEEE. 2023, pp. 1–5.

[96] Joan Greenbaum and Morten Kyng. *Design at work: Cooperative design of computer systems*. CRC Press, 2020.

[97] Robert Greif, Andrew Lockey, Jan Breckwoldt, Francesc Carmona, Patricia Conaghan, Artem Kuzovlev, Lucas Pflanzl-Knizacek, Ferenc Sari, Salma Shammet, Andrea Scapigliati, et al. "European Resuscitation Council Guidelines 2021: Education for resuscitation". In: *Resuscitation* 161 (2021), pp. 388–407.

[98] Julia Gruber, Dominik Stumpf, Bernhard Zapletal, Stephanie Neuhold, and Henrik Fischer. "Real-time feedback systems in CPR". In: *Trends in Anaesthesia and Critical Care* 2.6 (2012), pp. 287–294.

[99] Christina Gsaxner, Jianning Li, Antonio Pepe, Yuan Jin, Jens Kleesiek, Dieter Schmalstieg, and Jan Egger. "The HoloLens in medicine: A systematic review and taxonomy". In: *Medical Image Analysis* 85 (2023), p. 102757.

[100] Düsseldorf H-H-U. *G*Power*. Version 3.1.9.7. Release for Windows. 2025.

[101] Joon Ha, Priya Parekh, David Gamble, James Masters, Peter Jun, Thomas Hester, Timothy Daniels, and Mansur Halai. "Opportunities and challenges of using augmented reality and heads-up display in orthopaedic surgery: A narrative review". In: *Journal of clinical orthopaedics and trauma* 18 (2021), pp. 209–215.

[102] Thilo Hackert, Milena Sachsenmaier, Ulf Hinz, Lutz Schneider, Christoph W Michalski, Christoph Springfeld, Oliver Strobel, Dirk Jäger, Alexis Ulrich, and Markus W Büchler. "Locally advanced pancreatic cancer: neoadjuvant therapy with folfirinox results in resectability in 60% of the patients". In: *Annals of surgery* 264.3 (2016), pp. 457–463.

[103] Thilo Hackert, Oliver Strobel, Christoph W Michalski, André L Mihaljevic, Arianeb Mehrabi, Beat Müller-Stich, Christoph Berchtold, Alexis Ulrich, and Markus W Büchler. "The TRIANGLE operation–radical surgery after neoadjuvant treatment for advanced pancreatic cancer: a single arm observational study". In: *Hpb* 19.11 (2017), pp. 1001–1007.

[104] R Harari, A Altaweel, E Anderson, and CN Pozner. "AR Crisis Checklist Adherence in Simulated OR Emergencies". In: *JMIR XR* (2025). URL: https://xr.jmir.org/2025/1/e60792/.

[105] Sandra G Hart. "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908.

[106] Benjamin Hatscher and Christian Hansen. "Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain". In: *Proceedings of the 20th ACM international conference on multimodal interaction*. 2018, pp. 145–153.

[107] Morton L Heilig. "Sensorama simulator (1962)". In: *URL: http://www. freepatentsonline. com/3050870. html* 1 (2018).

[108] Julia Hertel, Sukran Karaosmanoglu, Susanne Schmidt, Julia Bräker, Martin Semmann, and Frank Steinicke. "A taxonomy of interaction techniques for immersive augmented reality based on an iterative literature review". In: *2021 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE. 2021, pp. 431–440.

[109] Jennifer Higa, Sonia Nkatha, Roxana Ramirez Herrera, Hani Marcus, Soojeong Yoo, Ann Blandford, and Jeremy Opie. "Augmented reality for endoscopic transsphenoidal surgery: evaluating design factors with neurosurgeons". In: *International Journal of Computer Assisted Radiology and Surgery* (2024), pp. 1–6.

[110] Emi Higashi, Keisuke Fukagawa, Rinka Kasimura, Yuima Kanamori, Akinori Minazuki, and Hidehiko Hayashi. "Development and evaluation of a corrective feedback system using augmented reality for the high-quality cardiopulmonary resuscitation training". In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2017, pp. 716–721.

[111] Jun Young Hong, Je Hyeok Oh, Chan Woong Kim, and Dong Hoon Lee. "Hand injuries caused by feedback device usage during cardiopulmonary resuscitation training". In: *Resuscitation* 107 (2016), e3–e4.

[112] Matthew B Hoy. "Alexa, Siri, Cortana, and more: an introduction to voice assistants". In: *Medical reference services quarterly* 37.1 (2018), pp. 81–88.

[113] Wonil Hwang and Gavriel Salvendy. "Number of people required for usability evaluation: the 10±2 rule". In: *Communications of the ACM* 53.5 (2010), pp. 130–133.

[114] Pier Luigi Ingrassia, Giulia Mormando, Eleonora Giudici, Francesco Strada, Fabio Carfagna, Fabrizio Lamberti, and Andrea Bottino. "Augmented reality learning environment for basic life support and defibrillation training: usability study". In: *Journal of medical Internet research* 22.5 (2020), e14910.

[115] C Innocente, L Ulrich, and S et al. Moos. "Augmented reality: mapping methods and tools for enhancing the human role in healthcare HMI". In: *Applied Sciences* (2022). URL: https://www.mdpi.com/2076-3417/12/9/4295.

[116] *Introduction to Mixed Reality Toolkit (MRTK) for Unity - MRTK2*. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05. Accessed: April 2024. Microsoft, 2023.

[117] Lorenzo Isella, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Ravà, Caterina Rizzo, et al. "Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors". In: *PloS one* 6.2 (2011), e17144.

[118] Ajune Wanis Ismail and Mohd Shahrizal Sunar. "Multimodal fusion: gesture and speech input in augmented reality environment". In: *Computational Intelligence in Information Systems: Proceedings of the Fourth INNS Symposia Series on Computational Intelligence in Information Systems (INNS-CIIS 2014)*. Springer. 2015, pp. 245–254.

[119] Malte Issleib, Alina Kromer, Hans O Pinnschmidt, Christoph Süss-Havemann, and Jens C Kubitz. "Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: a randomized controlled trial". In: *Scandinavian journal of trauma, resuscitation and emergency medicine* 29.1 (2021), pp. 1–9.

[120] K Ito, M Sugimoto, T Tsunoyama, and T Nagao. "A trauma patient care simulation using extended reality technology in the hybrid emergency room system". In: *Journal of Trauma and Acute Care Surgery* (2021). URL: https://journals.lww.com/jtrauma/fulltext/2021/05000/a_trauma_patient_care_simulation_using_extended.22.aspx.

[121] Vladimir M Ivanov, Anton M Krivtsov, Sergey V Strelkov, Anton Yu Smirnov, Roman Yu Shipov, Vladimir G Grebenkov, Valery N Rumyantsev, Igor S Gheleznyak, Dmitry A Surov, Michail S Korzhuk, et al. "Practical application of augmented/Mixed reality technologies in surgery of abdominal cancer patients". In: *Journal of Imaging* 8.7 (2022), p. 183.

[122] Lawrence R James, Robert G Demaree, and Gerrit Wolf. "Estimating within-group interrater reliability with and without response bias." In: *Journal of applied psychology* 69.1 (1984), p. 85.

[123] Hamraz Javaheri, Omid Ghamarnejad, Ragnar Bade, Paul Lukowicz, Jakob Karolus, and Gregor Alexander Stavrou. "Beyond the visible: preliminary evaluation of the first wearable augmented reality assistance system for pancreatic surgery". In: *International Journal of Computer Assisted Radiology and Surgery* (2024), pp. 1–13.

[124] Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. "LLMs Enable Context-Aware Augmented Reality in Surgical Navigation". In: *arXiv e-prints* (2024), arXiv–2412.

[125] Taoran Jiang, Dewang Yu, Yuqi Wang, Tao Zan, Shuyi Wang, and Qingfeng Li. "HoloLens-based vascular localization system: precision evaluation study with a three-dimensional printed model". In: *Journal of medical Internet research* 22.4 (2020), e16852.

[126] Xiangyan Jiang, Zeyuan Yu, Zhijian Ma, Haixiao Deng, Wen Ren, Wengui Shi, and Zuoyi Jiao. "Superior mesenteric artery first approach can improve the clinical outcomes of pancreaticoduodenectomy: a meta-analysis". In: *International Journal of Surgery* 73 (2020), pp. 14–24.

[127] Janet G Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel. "Holocpr: Designing and evaluating a mixed reality interface for time-critical emergencies". In: *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2018, pp. 67–76.

[128] L Jud, J Fotouhi, and O et al. Andronic. "Applicability of augmented reality in orthopedic surgery–a systematic review". In: *BMC Musculoskeletal Disorders* (2020). URL: https://link.springer.com/article/10.1186/s12891-020-3110-2.

[129] Christian Jung, Georg Wolff, Bernhard Wernly, Raphael Romano Bruno, Marcus Franz, P Christian Schulze, Jennifer N Avari Silva, Jonathan R Silva, Deepak L Bhatt, and Malte Kelm. "Virtual and augmented reality in cardiovascular care: state-of-the-art and future perspectives". In: *Cardiovascular Imaging* 15.3 (2022), pp. 519–532.

[130] Dominika Kanschik, Raphael Romano Bruno, Georg Wolff, Malte Kelm, and Christian Jung. "Virtual and augmented reality in intensive care medicine: a systematic review". In: *Annals of intensive care* 13.1 (2023), p. 81.

[131] Olga Kantor, Mark S Talamonti, Chi-Hsiung Wang, Kevin K Roggin, David J Bentrem, David J Winchester, Richard A Prinz, and Marshall S Baker. "The extent of vascular resection is associated with perioperative outcome in patients undergoing pancreaticoduodenectomy". In: *HPB* 20.2 (2018), pp. 140–146.

[132] William Kaye, Sharon F Rallis, Mary E Mancini, Kenneth C Linhares, Michael L Angell, Doris S Donovan, Nancy C Zajano, and John A Finger. "The problem of poor retention of cardiopulmonary resuscitation skills may lie with the instructor, not the learner or the curriculum". In: *Resuscitation* 21.1 (1991), pp. 67–87.

[133] M Kersten-Oertel, S Chen, and DL Collins. "Real-time multi-modal visualization for neurosurgical guidance using in-situ augmented reality". In: *Medical Image Analysis* (2020). URL: https://www.sciencedirect.com/science/article/pii/S136184152030084X.

[134] R Khan, A Alzahrani, M Aalsalem, and R Mehmood. "Healthcare data and applications in cloud: Applications, privacy, and security issues". In: *Future Generation Computer Systems* (2020). URL: https://www.sciencedirect.com/science/article/pii/S0167739X19324315.

[135] Chaowanan Khundam, Varunyu Vorachart, Patibut Preeyawongsakul, Witthaya Hosap, and Frédéric Noël. "A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training". In: *Informatics*. Vol. 8. 3. MDPI. 2021, p. 60.

[136] B Kim, YH Loke, P Mass, and MR et al. Irwin. "A novel virtual reality medical image display system for group discussions of congenital heart disease: development and usability testing". In: *JMIR Cardio* (2020). URL: https://cardio.jmir.org/2020/1/e20633/.

[137] H Kim, YT Kwon, HR Lim, and JH Kim. "Recent advances in wearable sensors and integrated functional devices for virtual and augmented reality applications". In: *Advanced Functional Materials* (2021). URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.202005692.

[138] H Kim, MS Zitouni, and A et al. Khandoker. "Trends in smart helmets with multimodal sensing for health and safety: scoping review". In: *JMIR mHealth and uHealth* (2022). URL: https://mhealth.jmir.org/2022/11/e40797.

[139] Hyun K Kim, Jaehyun Park, Yeongcheol Choi, and Mungyeong Choe. "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment". In: *Applied ergonomics* 69 (2018), pp. 66–73.

[140] Shelley Kirkbright, Judith Finn, Hideo Tohira, Alexandra Bremner, Ian Jacobs, and Antonio Celenza. "Audiovisual feedback device use by health care professionals during CPR: a systematic review and meta-analysis of randomised and non-randomised trials". In: *Resuscitation* 85.4 (2014), pp. 460–471.

[141] G Klein and D Murray. "Parallel tracking and mapping for small AR workspaces". In: *ISMAR* (2007). URL: https://ieeexplore.ieee.org/document/4538852.

[142] SN Kumar, AL Fred, and H Ajay Kumar. "Role of Augmented Reality and Virtual Reality in Medical Imaging". In: *Wiley* (2025). URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119857716.ch9.

[143] Katherine Kuyt, Sang-Hee Park, Todd P Chang, Timothy Jung, and Ralph MacKinnon. "The use of virtual reality and augmented reality to enhance cardio-pulmonary resuscitation: a scoping review". In: *Advances in Simulation* 6.1 (2021), pp. 1–8.

[144] Bambu Lab. *Bambu Lab P1S*. https://bambulab.com/en/p1?product=p1s. Accessed: 2025-01-20.

[145] Bambu Lab. *Bambu Studio*. https://bambulab.com/en/download/studio. Version 1.9.5.51. Accessed: 2025-01-20. 2025.

[146] Laerdal. *Little Anne CPR Dummy*. Official page. Retrieved July 06, 2018 from https://www.laerdal.com/de/. July 2018.

[147] V Lahanas, E Georgiou, and C Loukas. "Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators". In: *International Journal of Advanced Robotics and Automation* (2016). URL: https://www.academia.edu/download/113250940/robotics-automation09.pdf.

[148] Jessica A Latona, Sami Tannouri, Francesco Palazzo, and Michael J Pucci. "Fundamentals of Sutures, Needles, Knot Tying, and Suturing Technique". In: *Fundamentals of General Surgery* (2018), pp. 39–64.

[149] Marion Leary, Shaun K McGovern, Steve Balian, Benjamin S Abella, and Audrey L Blewer. "A pilot study of CPR quality comparing an augmented reality application vs. a standard audio-visual feedback manikin". In: *Frontiers in Digital Health* 2 (2020), p. 1.

[150] Vicki R LeBlanc. "The effects of acute stress on performance: implications for health professions education". In: *Academic Medicine* 84.10 (2009), S25–S33.

[151] Dongheon Lee, Jin Wook Yi, Jeeyoung Hong, Young Jun Chai, Hee Chan Kim, and Hyoun-Joong Kong. "Augmented reality to localize individual organ in surgical procedure". In: *Healthcare Informatics Research* 24.4 (2018), pp. 394–401.

[152] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. "A usability study of multimodal input in an augmented reality environment". In: *Virtual Reality* 17 (2013), pp. 293–305.

[153] Florian Leiser, Sven Eckhardt, Valentin Leuthe, Merlin Knaeble, Alexander Maedche, Gerhard Schwabe, and Ali Sunyaev. "Hill: A hallucination identifier for large language models". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–13.

[154] BSH Leong. "Bystander CPR and survival". In: *Singapore medical journal* 52.8 (2011), p. 573.

[155] Tal Levit, Patrick Grzela, Declan CT Lavoie, Li Wang, Aashna Agarwal, Rachel J Couban, and Harsha Shanthanna. "The effectiveness of virtual and augmented reality in surgical pain management: A systematic review of randomized controlled trials". In: *Anesthesia & Analgesia* (2022), pp. 10–1213.

[156] James R Lewis, Brian S Utesch, and Deborah E Maher. "Investigating the correspondence between UMUX-LITE and SUS scores". In: *Design, User Experience, and Usability: Design Discourse: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings, Part I*. Springer. 2015, pp. 204–211.

[157] James R Lewis, Brian S Utesch, and Deborah E Maher. "UMUX-LITE: when there's no time for the SUS". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013, pp. 2099–2102.

[158] CHJ Li, V Liang, and YTH et al. Chow. "A mixed reality-based platform towards human-cyber-physical systems with IoT wearable device for occupational safety and health training". In: *Applied Sciences* (2022). URL: https://www.mdpi.com/2076-3417/12/23/12009.

[159] Debra K Lindstrom-Hazel and Nicole VanderVlies Veenstra. "Examining the Purdue pegboard test for occupational therapy practice". In: *The Open Journal of Occupational Therapy* 3.3 (2015), p. 5.

[160] Kai Liu, Yuan Gao, Ahmed Abdelrehem, Lei Zhang, Xi Chen, Le Xie, and Xudong Wang. "Augmented reality navigation method for recontouring surgery of craniofacial fibrous dysplasia". In: *Scientific Reports* 11.1 (2021), p. 10043.

[161] RG Louis, GK Steinberg, and C et al. Duma. "Early experience with virtual and synchronized augmented reality platform for preoperative planning and intraoperative navigation: a case series". In: *Operative Neurosurgery* (2021). URL: https://journals.lww.com/onsonline/fulltext/2021/10000/early_experience_with_virtual_and_synchronized.3.aspx.

[162] AJ Lungu, W Swinkels, L Claesen, and P Tu. "A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation". In: *Expert Review of Medical Devices* (2021). URL: https://www.tandfonline.com/doi/abs/10.1080/17434440.2021.1860750.

[163] Weizhou Luo, Eva Goebel, Patrick Reipschläger, Mats Ole Ellenberg, and Raimund Dachselt. "Exploring and slicing volumetric medical data in augmented reality using a spatially-aware mobile device". In: *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2021, pp. 334–339.

[164] Longfei Ma, Tianqi Huang, Jie Wang, and Hongen Liao. "Visualization, registration and tracking techniques for augmented reality guided surgery: a review". In: *Physics in Medicine & Biology* 68.4 (2023), 04TR02.

[165] Amama Mahmood, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-Ming Huang. "LLM-Powered Conversational Voice Assistants: Interaction Patterns, Opportunities, Challenges, and Design Guidelines". In: *arXiv preprint arXiv:2309.13879* (2023).

[166] L Maier-Hein, P Mountney, and A et al. Bartoli. "Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery". In: *Medical Image Analysis* (2013). URL: https://www.sciencedirect.com/science/article/pii/S1361841513000740.

[167] HA Maqsood, M Farooq, L Sarfraz, and M Truitt. "Telestration with Augmented Reality in Surgery: A Comprehensive Systematic Review". In: *HCA Healthcare Scholarly Commons* (2025).

[168] Ettore Marzano, Tullio Piardi, Luc Soler, Michele Diana, Didier Mutter, Jacques Marescaux, and Patrick Pessaux. "Augmented reality-guided artery-first pancreatico-duodenectomy". In: *Journal of Gastrointestinal Surgery* 17 (2013), pp. 1980–1983.

[169] Mouayad Masalkhi, Ethan Waisberg, Joshua Ong, Nasif Zaman, Prithul Sarker, Andrew G Lee, and Alireza Tavakkoli. "Apple vision pro for ophthalmology and medicine". In: *Annals of Biomedical Engineering* 51.12 (2023), pp. 2643–2646.

[170] C Mather and T Barnett. "Helping Hands: Using AR to Provide Remote Guidance to Health Professionals". In: *IOS Press* (2017). URL: https://ebooks.iospress.nl/pdf/doi/10.3233/978-1-61499-794-8-57.

[171] Nadine S Matthie, Nicholas A Giordano, Coretta M Jenerette, Gayenell S Magwood, Sharon L Leslie, Emily E Northey, Caitlin I Webster, and Soumitri Sil. "Use and efficacy of virtual, augmented, or mixed reality technology for chronic pain: a systematic review". In: *Pain Management* 12.7 (2022), pp. 859–878.

[172] Calvin R Maurer, J Michael Fitzpatrick, Matthew Y Wang, Robert L Galloway, Robert J Maciunas, and George S Allen. "Registration of head volume images using implantable fiducial markers". In: *IEEE transactions on medical imaging* 16.4 (1997), pp. 447–462.

[173] E McMullen, S Kamurari, R Price, and M Mulimira. "Emerging Technology Solutions to Support Emergency Workforce Training in Uganda". In: *Emergency Medicine Journal* (2025). URL: https://emj.bmj.com/content/emermed/42/1/62.full.pdf.

[174] Mediarchitect. *Suture Practice*. 2025. URL: https://mediarchitect.net/en-de/collections/suture-practice.

[175] Laerdal Medical. *CPRmeter 2*. Accessed: Spetember 2023. Year. URL: https://laerdal.com/de/products/medical-devices/cpr-feedback-devices/cprmeter-2/.

[176] Helena M Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. "Voice or gesture in the operating room". In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 2015, pp. 773–780.

[177] Meta Platforms Inc. *Introducing Meta Quest Pro: A New Era for Mixed Reality*. Website. Accessed: 2025-03-15. 2025. URL: https://forwork.meta.com/de/meta-horizon-managed-solutions/quest-3/.

[178] MeVis Medical Solutions AG. *MeVis Liver Software*. https://www.mevis.de/en/. Accessed: 2024-03-06.

[179] Microsoft. *HoloLens*. Official page. Retrieved July 06, 2018 from https://www.microsoft.com/en-us/hololens. July 2018.

[180] Microsoft. *Microsoft HoloLens 2*. https://www.microsoft.com/en-us/hololens. Accessed: 2024-03-06. 2023.

[181] Andre L Mihaljevic, Thilo Hackert, Martin Loos, Ulf Hinz, Martin Schneider, Arianeb Mehrabi, Katrin Hoffmann, Christoph Berchtold, Beat P Müller-Stich, Markus Diener, et al. "Not all Whipple procedures are equal: proposal for a classification of pancreatoduodenectomies". In: *Surgery* 169.6 (2021), pp. 1456–1462.

[182] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. "Augmented reality: A class of displays on the reality-virtuality continuum". In: *Telemanipulator and telepresence technologies*. Vol. 2351. Spie. 1995, pp. 282–292.

[183] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The ethics of algorithms: Mapping the debate". In: *Big Data & Society* 3.2 (2016), p. 2053951716679679.

[184] TB Moeslund, A Hilton, and V Krüger. "A survey of advances in vision-based human motion capture and analysis". In: *Computer Vision and Image Understanding* (2011). URL: https://www.sciencedirect.com/science/article/pii/S1077314210001360.

[185] Christian Moro, Charlotte Phelps, Petrea Redmond, and Zane Stromberga. "HoloLens and mobile augmented reality in medical and health science education: A randomised controlled trial". In: *British Journal of Educational Technology* 52.2 (2021), pp. 680–694.

[186] Peter Mountney and Guang-Zhong Yang. "Motion compensated SLAM for image guided surgery". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part II 13*. Springer. 2010, pp. 496–504.

[187] Brendan William Munzer, Mohammad Mairaj Khan, Barbara Shipman, and Prashant Mahajan. "Augmented reality in emergency medicine: a scoping review". In: *Journal of medical Internet research* 21.4 (2019), e12368.

[188] N Navab, T Blum, A Okur, and T Wendler. "First deployments of augmented reality in operating rooms". In: *Computer* (2015). URL: `https://ieeexplore.ieee.org/document/7139473`.

[189] N Navab, SM Heining, and J Traub. "Camera augmented mobile C-arm (CAMC): Calibration, accuracy study, and clinical applications". In: *IEEE Transactions on Medical Imaging* (2004). URL: `https://ieeexplore.ieee.org/document/1320553`.

[190] MD Negrão. "Characterizing affordances using AR for laparoscopic surgeries and challenges of planning a mixed reality remote collaboration application". In: *University of Rio Grande do Sul Repository* (2024). URL: `https://lume.ufrgs.br/handle/10183/273210`.

[191] Trong Nguyen, William Plishker, Andrew Matisoff, Karun Sharma, and Raj Shekhar. "HoloUS: Augmented reality visualization of live ultrasound images using HoloLens for ultrasound-guided procedures". In: *International journal of computer assisted radiology and surgery* 17.2 (2022), pp. 385–391.

[192] SS Muhammad Nizam, Rimaniza Zainal Abidin, Nurhazarifah Che Hashim, Meng Chun Lam, Haslina Arshad, and NAA Majid. "A review of multimodal interaction technique in augmented reality environment". In: *Int. J. Adv. Sci. Eng. Inf. Technol* 8.4-2 (2018), p. 1460.

[193] A Norman Donald. *The design of everyday things*. MIT Press, 2013.

[194] A Ojieh, S Mettu, J Balaji, and R Warren. "Virtual Reality as a Teaching Aid for Medical Professionals: A Transformative Approach to Healthcare Training". In: *ISTES Proceedings* (2024). URL: `https://www.istes.org/storage/01JFQJBQMGRSPA1C8SGEEY2VHT.pdf`.

[195] Tomoyoshi Okamoto, Shinji Onda, Jungo Yasuda, Katsuhiko Yanaga, Naoki Suzuki, and Asaki Hattori. "Navigation surgery using an augmented reality for pancreatectomy". In: *Digestive surgery* 32.2 (2015), pp. 117–123.

[196] Theresa M Olasveengen, Federico Semeraro, Giuseppe Ristagno, Maaret Castren, Anthony Handley, Artem Kuzovlev, Koenraad G Monsieurs, Violetta Raffay, Michael Smyth, Jasmeet Soar, et al. "European resuscitation council guidelines 2021: basic life support". In: *Resuscitation* 161 (2021), pp. 98–114.

[197] Karl J Oldhafer, Gregor A Stavrou, Guido Prause, Heinz-Otto Peitgen, Tim C Lueth, and Stefan Weber. "How to operate a liver tumor you cannot see". In: *Langenbeck's archives of surgery* 394 (2009), pp. 489–494.

[198] Joshua Olexa, Annie Trang, Jonathan Cohen, Kevin Kim, Maureen Rakovec, Jordan Saadon, Charles Sansur, Graeme Woodworth, Gary Schwartzbauer, Jacob Cherian, et al. "The Apple Vision Pro as a neurosurgical planning tool: a case report". In: *Cureus* 16.2 (2024).

[199] Shinji Onda, Tomoyoshi Okamoto, Masaru Kanehira, Shuichi Fujioka, Naoki Suzuki, Asaki Hattori, and Katsuhiko Yanaga. "Short rigid scope and stereo-scope designed specifically for open abdominal navigation surgery: clinical application for hepatobiliary and pancreatic surgery". In: *Journal of hepato-biliary-pancreatic sciences* 20 (2013), pp. 448–453.

[200] Shinji Onda, Tomoyoshi Okamoto, Masaru Kanehira, Fumitake Suzuki, Ryusuke Ito, Shuichi Fujioka, Naoki Suzuki, Asaki Hattori, and Katsuhiko Yanaga. "Identification of inferior pancreaticoduodenal artery during pancreaticoduodenectomy using augmented reality-based navigation system". In: *Journal of Hepato-Biliary-Pancreatic Sciences* 21.4 (2014), pp. 281–287.

[201] M Orione, G Rubegni, R Tartaro, A Alberghina, M Fallico, C Orione, A Russo, GM Tosi, and T Avitabile. "Utilization of apple vision pro in ophthalmic surgery: A pilot study". In: *European Journal of Ophthalmology* (2024), p. 11206721241273574.

[202] A Oun, N Hagerdorn, and X Cheng. "Mobile Devices or Head-Mounted Displays: A Comparative Review and Analysis of Augmented Reality in Healthcare". In: *IEEE Access* (2024). URL: `https://ieeexplore.ieee.org/document/10419075`.

[203] Martina Paglino, Enrico Contri, Marta Baggiani, Michela Tonani, Giulia Costantini, Maria Concetta Bonomo, and Enrico Baldi. "A video-based training to effectively teach CPR with long-term retention: the ScuolaSalvaVita. it ("SchoolSavesLives. it") project". In: *Internal and emergency medicine* 14.2 (2019), pp. 275–279.

[204] "Pancreatic surgery outcomes: multicentre prospective snapshot study in 67 countries". In: *British Journal of Surgery* 111.1 (2024), znad330.

[205] T Papadopoulos, K Evangelidis, and TH Kaskalis. "Interactions in augmented and mixed reality: an overview". In: *Applied Sciences* (2021). URL: `https://www.mdpi.com/2076-3417/11/18/8752`.

[206] Vaidehi Patil, Jyotindra Narayan, Kamalpreet Sandhu, and Santosha K Dwivedy. "Integration of virtual reality and augmented reality in physical rehabilitation: a state-of-the-art review". In: *Revolutions in Product Design for Healthcare: Advances in Product Design and Design Methods for Healthcare* (2022), pp. 177–205.

[207] Long Peng, Shengrong Lin, Yong Li, and Weidong Xiao. "Systematic review and meta-analysis of robotic versus open pancreaticoduodenectomy". In: *Surgical endoscopy* 31 (2017), pp. 3085–3097.

[208] Gavin D Perkins, Jan-Thorsen Gräsner, Federico Semeraro, Theresa Olasveengen, Jasmeet Soar, Carsten Lott, Patrick Van de Voorde, John Madar, David Zideman, Spyridon Mentzelopoulos, et al. "European Resuscitation Council Guidelines 2021: Executive summary". In: *Resuscitation* 161 (2021), pp. 1–60.

[209] Stephanie L Perkins, Michael A Lin, Subashini Srinivasan, Amanda J Wheeler, Brian A Hargreaves, and Bruce L Daniel. "A mixed-reality system for breast surgical planning". In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE. 2017, pp. 269–274.

[210] T Piumsomboon, G Lee, A Clark, and M Billinghurst. "Exploring natural eye-gaze-based interaction for immersive virtual reality". In: *IEEE Symposium on 3D User Interfaces* (2017). URL: https://ieeexplore.ieee.org/document/7893313.

[211] A Plopski, T Hirzle, N Norouzi, and L Qian. "The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality". In: *ACM Computing Surveys* (2022). URL: https://dl.acm.org/doi/abs/10.1145/3491207.

[212] Vitaliy Popov, Xinyue Chen, Jingying Wang, Michael Kemp, Gurjit Sandhu, Taylor Kantor, Natalie Mateju, and Xu Wang. "Looking Together≠ Seeing the Same Thing: Understanding Surgeons' Visual Needs During Intra-operative Coordination and Instruction". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–12.

[213] PTC Inc. *Vuforia Augmented Reality*. Accessed: 2024-05-08. 2024. URL: https://developer.vuforia.com/.

[214] AP Putro and MK Biddinika. "Implementasi Model 3D Jantung Manusia pada Aplikasi Markerless Augmented Reality". In: *Jurnal JTIK* (2025). URL: https://journal.lembagakita.org/jtik/article/view/3657.

[215] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: https://www.R-project.org/.

[216] R Core Team. *R: A Language and Environment for Statistical Computing*. Version 4.5.0. R Foundation for Statistical Computing. Vienna, Austria, 2025. URL: https://www.R-project.org/.

[217] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2025. URL: https://www.r-project.org/.

[218] João Ramalhinho, Soojeong Yoo, Thomas Dowrick, Bongjin Koo, Murali Somasundaram, Kurinchi Gurusamy, David J Hawkes, Brian Davidson, Ann Blandford, and Matthew J Clarkson. "The value of Augmented Reality in surgery—A usability study on laparoscopic liver surgery". In: *Medical Image Analysis* 90 (2023), p. 102943.

[219] Isabelle Rapin, Lester M Tourk, and Louis D Costa. "Evaluation of the Purdue Pegboard as a screening test for brain damage". In: *Developmental Medicine & Child Neurology* 8.1 (1966), pp. 45–54.

[220] Anke V Reinschluessel and Johannes Zagermann. "Exploring Hybrid User Interfaces for Surgery Planning". In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2023, pp. 208–210.

[221] Anke Verena Reinschluessel, Thomas Muender, Roland Fischer, Valentin Kraft, Verena Nicole Uslar, Dirk Weyhe, Andrea Schenk, Gabriel Zachmann, Tanja Döring, and Rainer Malaka. "Versatile immersive virtual and augmented tangible or–using vr, ar and tangibles to support surgical practice". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–5.

[222] Serena Ricci, Andrea Calandrino, Giacomo Borgonovo, Marco Chirico, and Maura Casadio. "Virtual and augmented reality in basic and advanced life support training". In: *JMIR Serious Games* 10.1 (2022), e28595.

[223] PG Rodrigues, DMK Leandro, and SS et al. de Azevedo. "Transforming neonatal care: a position paper on the potential of augmented and mixed reality". In: *Frontiers in Digital Health* (2025). URL: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.1571521.

[224] R Romero-Reveron. "Changes Times in Teaching / Learning Human Anatomy". In: *WebLog J Anat* (2025). URL: https://www.researchgate.net/publication/391344995.

[225] Yu Saito, Maki Sugimoto, Satoru Imura, Yuji Morine, Tetsuya Ikemoto, Shuichi Iwahashi, Shinichiro Yamada, and Mitsuo Shimada. "Intraoperative 3D hologram support with mixed reality techniques in liver surgery". In: *Annals of surgery* 271.1 (2020), e4–e7.

[226] Golnoosh Samei, Keith Tsang, Claudia Kesch, Julio Lobo, Soheil Hor, Omid Mohareri, Silvia Chang, S Larry Goldenberg, Peter C Black, and Septimiu Salcudean. "A partial augmented reality system with live ultrasound and registered preoperative MRI for guiding robot-assisted radical prostatectomy". In: *Medical image analysis* 60 (2020), p. 101588.

[227] TBM Sampaio, ACM do Rêgo, and I Araújo-Filho. "Development of a Platform in Augmented Reality for Remote Surgical Training in the Context of Health Emergencies and Environmental Disasters". In: *ResearchGate* (2025). URL: https://www.researchgate.net/publication/%20390944003.

[228] Ezequiel Santos and Vanessa Castillo. "Apple vision pro: Comments in healthcare". In: *arXiv preprint arXiv:2401.08685* (2024).

[229] Claudia Scherl, David Männle, Nicole Rotter, Jürgen Hesser, Jan Stallkamp, Tobias Balkenhol, Lena Huber, Benedikt Kramer, Anne Lammert, and Annette Affolter. "Augmented reality during parotid surgery: real-life evaluation of voice control of a head mounted display". In: *European Archives of Oto-Rhino-Laryngology* 280.4 (2023), pp. 2043–2049.

[230] J Schild and C Elsenbast. "ViTAWiN: Developing Multiprofessional Medical Emergency Training with Mixed Reality". In: *IEEE Conference on Healthcare Applications* (2021). URL: https://www.researchgate.net/publication/355138869_ViTAWiN.

[231] Masaki Sekine, Toshiyo Tamura, Masaki Yoshida, Yuki Suda, Yuichi Kimura, Hiroaki Miyoshi, Yoshifumi Kijima, Yuji Higashi, and Toshiro Fujimoto. "A gait abnormality measure based on root mean square of trunk acceleration". In: *Journal of neuroengineering and rehabilitation* 10 (2013), pp. 1–7.

[232] David P de Sena, Daniela D Fabrício, Vinícius D da Silva, Luiz Carlos Bodanese, and Alexandre R Franco. "Comparative evaluation of video-based on-line course versus serious game for training medical students in cardiopulmonary resuscitation: a randomised trial". In: *PloS one* 14.4 (2019), e0214722.

[233] Nandan M Shanbhag, Abdulrahman Bin Sumaida, Khalifa Al Shamisi, and Khalid Balaraj. "Apple Vision Pro: a paradigm shift in medical technology". In: *Cureus* 16.9 (2024).

[234] Y Shen, S Wang, Y Shen, and J Hu. "The Application of Augmented Reality Technology in Perioperative Visual Guidance: Technological Advances and Innovation Challenges". In: *Sensors* (2024). URL: https://www.mdpi.com/1424-8220/24/22/7363.

[235] Ben Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations". In: *The craft of information visualization.* Elsevier, 2003, pp. 364–371.

[236] T Sielhorst, M Feuerstein, and N Navab. "Advanced medical displays: A literature review of augmented reality". In: *Journal of Display Technology* (2008). URL: https://ieeexplore.ieee.org/document/4479595.

[237] Jesper Simonsen and Toni Robertson. *Routledge international handbook of participatory design.* Vol. 711. Routledge New York, 2013.

[238] Jingwei Song, Jun Wang, Liang Zhao, Shoudong Huang, and Gamini Dissanayake. "Mis-slam: Real-time large-scale dense deformable slam system in minimal invasive surgery based on heterogeneous computing". In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 4068–4075.

[239] Tianyu Song, Kevin Yu, Ulrich Eck, and Nassir Navab. "Augmented reality collaborative medical displays (ARC-MeDs) for multi-user surgical planning and intra-operative communication". In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 11.4 (2023), pp. 1042–1049.

[240] M Sorani. "AI-AR Integrated Emergency Preparedness Training in Trauma Care". In: *JPCM* (2025). URL: `https://www.jpcmed.com/article_215785.html`.

[241] Otfried Spreen and Esther Strauss. *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press, 1998.

[242] Salaheddine Sta, Jérôme Ogor, Hoel Letissier, Eric Stindel, Chafiaa Hamitouche, and Guillaume Dardenne. "Towards markerless computer assisted surgery: Application to total knee arthroplasty". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 17.5 (2021), e2296.

[243] Oliver Strobel, Thomas Hank, Ulf Hinz, Frank Bergmann, Lutz Schneider, Christoph Springfeld, Dirk Jäger, Peter Schirmacher, Thilo Hackert, and Markus W Büchler. "Pancreatic cancer surgery: the new R-status counts". In: *Annals of surgery* 265.3 (2017), pp. 565–573.

[244] Lucille Alice Suchman. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.

[245] Sungho Suh, Vitor Fortes Rey, and Paul Lukowicz. "Tasked: transformer-based adversarial learning for human activity recognition using wearable sensors via self-knowledge distillation". In: *Knowledge-Based Systems* 260 (2023), p. 110143.

[246] Sungho Suh, Vitor Fortes Rey, and Paul Lukowicz. "Wearable Sensor-Based Human Activity Recognition for Worker Safety in Manufacturing Line". In: *Artificial Intelligence in Manufacturing: Enabling Intelligent, Flexible and Cost-Effective Production Through AI*. Springer, 2023, pp. 303–317.

[247] B Šumak, S Brdnik, and M Pušnik. "Sensors and AI methods for human–computer intelligent interaction: A systematic mapping study". In: *Sensors* (2021). URL: `https://www.mdpi.com/1424-8220/22/1/20`.

[248] P Sun, Y Zhao, and J et al. Men. "Application of virtual and augmented reality technology in hip surgery: systematic review". In: *Journal of Medical Internet Research* (2023). URL: `https://www.jmir.org/2023/1/e37599/`.

[249] D Suresh, A Aydin, S James, and K Ahmed. "The role of augmented reality in surgical training: a systematic review". In: *Surgical Innovation* (2023). URL: https://journals.sagepub.com/doi/abs/10.1177/15533506221140506.

[250] BA Taha, AC Kadhim, and AS et al. Azzahran. "Photonics-powered augmented reality skin electronics for proactive healthcare: multifaceted opportunities". In: *Microchimica Acta* (2024). URL: https://link.springer.com/article/10.1007/s00604-024-06314-3.

[251] RR Tamboli and VK et al. Adhikarla. "Connected without Disconnection: Overview of Light Field Metaverse Applications and Their Quality of Experience". In: *Displays* (2023). URL: https://www.sciencedirect.com/science/article/pii/S014193822300063X.

[252] Rui Tang, Wei Yang, Yucheng Hou, Lihan Yu, Guangdong Wu, Xuan Tong, Jun Yan, and Qian Lu. "Augmented reality-assisted pancreaticoduodenectomy with superior mesenteric vein resection and reconstruction". In: *Gastroenterology research and practice* 2021 (2021), pp. 1–7.

[253] George Terzopoulos and Maya Satratzemi. "Voice assistants and smart speakers in everyday life and in education". In: *Informatics in Education* 19.3 (2020), pp. 473–490.

[254] Joseph Tiffin and ESTON J Asher. "The Purdue Pegboard: norms and studies of reliability and validity." In: *Journal of applied psychology* 32.3 (1948), p. 234.

[255] Yu-Chieh Tsai, Chin-Cheng Hsiao, Charles Chung-Wei Lin, Ming-Chih Chern, and Shih-Wei Huang. "Apple Vision Pro-guided Laparoscopic Radio Frequency Ablation for Liver Tumors: The Pioneer Experience". In: *Surgical Innovation* (2025), p. 15533506251316001.

[256] Jui-Yi Tsou, Chia-Lung Kao, Ming-Yuan Hong, Chih-Jan Chang, Fong-Chin Su, and Chih-Hsien Chi. "How does the side of approach impact the force delivered during external chest compression?" In: *The American Journal of Emergency Medicine* 48 (2021), pp. 67–72.

[257] N Tueni and F Amirouche. "Branding a New Technological Outlook for Future Orthopaedics". In: *Bioengineering* (2025). URL: https://www.mdpi.com/2306-5354/12/5/494.

[258] Unity. *Unity 3D Game Engine 2020.3.X*. https://unity.com/. Accessed: 2024-03-06. 2024.

[259] Julie Urlings, Sümeyye Sezer, Mark Ter Laan, Ronald Bartels, Thomas Maal, Jeroen Boogaarts, and Dylan Henssen. "The role and effectiveness of augmented reality in patient education: a systematic review of the literature". In: *Patient Education and Counseling* 105.7 (2022), pp. 1917–1927.

[260] Basil Varkey. "Principles of clinical ethics and their application to practice". In: *Medical Principles and Practice* 30.1 (2021), pp. 17–28.

[261] P Vávra, J Roman, P Zonča, and P Ihnát. "Recent development of augmented reality in surgery: a review". In: *Journal of Healthcare Engineering* (2017). URL: https://onlinelibrary.wiley.com/doi/abs/10.1155/2017/4574172.

[262] Petr Vávra, Jan Roman, Pavel Zonča, Peter Ihnát, Martin Němec, Jayant Kumar, Nagy Habib, Ahmed El-Gendi, et al. "Recent development of augmented reality in surgery: a review". In: *Journal of healthcare engineering* 2017 (2017).

[263] Donald Venes. *Taber's cyclopedic medical dictionary*. FA Davis, 2017.

[264] RM Viglialoro, N Esposito, and S Condino. "Augmented reality to improve surgical simulation: Lessons learned towards the design of a hybrid laparoscopic simulator for cholecystectomy". In: *IEEE Transactions* (2018). URL: https://ieeexplore.ieee.org/abstract/document/8550677/.

[265] Laurent Voisard, Amal Hatira, Mohammad Raihanul Bashar, Mucahit Gemici, Mine Sarac, Marta Kereten-Oertel, and Anil Ufuk Batmaz. "Subtask-Based Virtual Hand Visualization Method for Enhanced User Accuracy in Virtual Reality Environments". In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE. 2024, pp. 6–11.

[266] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. "Vision-based hand-gesture applications". In: *Communications of the ACM* 54.2 (2011), pp. 60–71.

[267] Ethan Waisberg, Joshua Ong, Mouayad Masalkhi, Nasif Zaman, Prithul Sarker, Andrew G Lee, and Alireza Tavakkoli. "Apple Vision Pro and why extended reality will revolutionize the future of medicine". In: *Irish Journal of Medical Science (1971-)* 193.1 (2024), pp. 531–532.

[268] Gregory K Wanner, Arayel Osborne, and Charlotte H Greene. "Brief compression-only cardiopulmonary resuscitation training video and simulation with homemade mannequin improves CPR skills". In: *BMC emergency medicine* 16.1 (2016), pp. 1–6.

[269] Rong Wen, Chin-Boon Chng, and Chee-Kong Chui. "Augmented reality guidance with multimodality imaging data and depth-perceived interaction for robot-assisted surgery". In: *Robotics* 6.2 (2017), p. 13.

[270] MA Williams, J McVeigh, and AI Handa. "Augmented reality in surgical training: a systematic review". In: *Postgraduate Medical Journal* (2020). URL: https://academic.oup.com/pmj/article-abstract/96/1139/537/6958936.

[271] *WindowsSpeechInputProvider Class - Mixed Reality Toolkit*. https://learn.microsoft.com/en-us/dotnet/api/microsoft.mixedreality.toolkit.windows.input.windowsspeechinputprovider?preserve-view=true&view=mixed-reality-toolkit-unity-2020-dotnet-2.8.0. Accessed: April 2024. Microsoft, 2020.

[272] Bob G Witmer and Michael J Singer. "Measuring presence in virtual environments: A presence questionnaire". In: *Presence* 7.3 (1998), pp. 225–240.

[273] Haibin Wu, Ruotong Xu, Kaiyang Xu, Jianbo Zhao, Yan Zhang, Aili Wang, and Yuji Iwahori. "3D texture reconstruction of abdominal cavity based on monocular vision SLAM for minimally invasive surgery". In: *Symmetry* 14.2 (2022), p. 185.

[274] YH Andrew Wu, Atsushi Oba, Ronggui Lin, Shuichi Watanabe, Cheryl Meguid, Richard D Schulick, and Marco Del Chiaro. "Selecting surgical candidates with locally advanced pancreatic cancer: a review for modern pancreatology". In: *Journal of Gastrointestinal Oncology* 12.5 (2021), p. 2475.

[275] L Yang, A Al-Ani, MS Bondok, P Gooi, and H Chung. "The Impact of Extended Reality Simulators on Ophthalmic Surgical Training". In: *Eye (Nature)* (2025). URL: https://www.nature.com/articles/s41433-025-03722-4.

[276] Jia-Yu Yao, Kun-Peng Ning, Zhen-Hui Liu, Mu-Nan Ning, and Li Yuan. "Llm lies: Hallucinations are not bugs, but features as adversarial examples". In: *arXiv preprint arXiv:2310.01469* (2023).

[277] Andy Wai Kan Yeung, Anela Tosevska, Elisabeth Klager, Fabian Eibensteiner, Daniel Laxar, Jivko Stoyanov, Marija Glisic, Sebastian Zeiner, Stefan Tino Kulnik, Rik Crutzen, et al. "Virtual and augmented reality applications in medicine: analysis of the scientific literature". In: *Journal of medical internet research* 23.2 (2021), e25499.

[278] Joyce Yeung, Reylon Meeks, Dana Edelson, Fang Gao, Jasmeet Soar, and Gavin D Perkins. "The use of CPR feedback/prompt devices during training and CPR performance: a systematic review". In: *Resuscitation* 80.7 (2009), pp. 743–751.

[279] Enes Yigitbas, Sebastian Krois, Timo Renzelmann, and Gregor Engels. "Comparative Evaluation of AR-based, VR-based, and Traditional Basic Life Support Training". In: *2022 IEEE 10th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE. 2022, pp. 1–8.

[280] Jang W Yoon, Robert E Chen, Esther J Kim, Oluwaseun O Akinduro, Panagiotis Kerezoudis, Phillip K Han, Phong Si, William D Freeman, Roberto J Diaz, Ricardo J Komotar, et al. "Augmented reality for the surgeon: systematic review". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 14.4 (2018), e1914.

[281] S Zhang, L Suresh, J Yang, and X Zhang. "Augmenting sensor performance with machine learning towards smart wearable sensing electronic systems". In: *Advanced Intelligent Systems* (2022). URL: https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202100194.

[282] XC Zhang, L Kobayashi, and SA Collins. "Exploratory application of augmented reality/mixed reality devices for acute care procedure training". In: *Western Journal of Emergency Medicine* (2017). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC5785186/.

[283] G Zheng, X Dong, and M Staring. "Registration of point clouds in surgical navigation: A review of methods and applications". In: *Computer Methods in Biomechanics and Biomedical Engineering* (2016). URL: https://www.tandfonline.com/doi/abs/10.1080/10255842.2015.1069580.

[284] Jiangjiao Zhou, Zhuo Lv, Heng Zou, LI Xiong, Zhongtao Liu, Wenhao Chen, and Yu Wen. "Up-to-date comparison of robotic-assisted versus open distal pancreatectomy: a PRISMA-compliant meta-analysis". In: *Medicine* 99.23 (2020), e20435.

Appendix A

Glossary of Medical Terms

Neoadjuvant chemotherapy Chemotherapy administered before surgery to shrink a tumor and increase the chances of complete removal.

Hepatopancreatobiliary (HPB) A medical field concerning diseases of the liver (hepato-), pancreas (pancreato-), and biliary system (bile ducts and gallbladder).

Retroperitoneum A space in the abdominal cavity behind the peritoneum. Organs such as the pancreas, kidneys, and aorta are located here.

R0 resection A complete surgical removal of a tumor with no cancer cells at the microscopic level along the resection margins.

R1 resection Tumor removal where cancer cells are present at the resection margin under the microscope, indicating incomplete excision.

Clavien-Dindo classification A system to grade the severity of surgical complications based on the type of treatment needed, ranging from minor (Grade I) to fatal (Grade V).

Postoperative pancreatic fistula (POPF) Leakage of pancreatic fluid from the surgical site; classified by severity using ISGPS definitions.

Delayed gastric emptying (DGE) A condition where the stomach fails to empty normally after surgery, leading to delayed oral intake and prolonged recovery.

Postpancreatectomy hemorrhage (PPH) Bleeding that occurs after pancreatic surgery. It is categorized into grades A, B, or C depending on severity and intervention required.

Appleby procedure A radical surgical technique involving removal of the distal pancreas and the celiac artery, typically used for tumors involving major arteries.

Borderline resectable pancreatic cancer Tumors that involve nearby major blood vessels. Surgery is possible but challenging; often treated with preoperative chemotherapy to improve outcomes.

ISGPS International Study Group of Pancreatic Surgery: An expert panel that standardizes definitions of complications related to pancreatic surgery.

PPPD (Pylorus-preserving pancreaticoduodenectomy) A type of surgery that removes the head of the pancreas while preserving the stomach's pylorus (outlet valve), commonly used for pancreatic cancer.

ASA classification A grading system developed by the American Society of Anesthesiologists to assess the preoperative health status of a patient.

Vascular resection Surgical removal of blood vessels that are invaded by a tumor; often involves reconstruction of the vessel during surgery.

Extended organ resection Surgical removal of additional organs (e.g., colon, stomach, liver) that are directly involved or invaded by the tumor.

Resection margin The edge of the tissue removed during surgery. Analysis determines whether the tumor was entirely excised (R0) or not (R1/R2).

Superior mesenteric artery (SMA) A major artery that supplies blood to the small intestine. Its proximity to the pancreas makes it a critical structure in pancreatic surgery.

Celiac axis A major artery branching off the aorta, supplying blood to the liver, stomach, and spleen. Often involved in advanced pancreatic cancers.

List of Publications

1. Javaheri, Hamraz, et al. "When AR Hinders Performance: The Hidden Costs of Video-See-Through Displays." Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology. 2025.

2. Javaheri, Hamraz, Vitor Fortes Rey, Paul Lukowicz, Gregor A. Stavrou, Jakob Karolus, Omid Ghamarnejad. "Assessing the Feasibility of Using Apple Vision Pro While Performing Medical Precision Tasks: Controlled User Study." *JMIR XR Spatial Comput*, 2025;2:e73574. doi: 10.2196/73574.

3. Javaheri, Hamraz, et al. "RescuAR: A Self-Directed Augmented Reality System for Cardiopulmonary Resuscitation Training." *International Conference on Pervasive Computing Technologies for Healthcare*. Cham: Springer Nature Switzerland, 2023.

4. Javaheri, Hamraz, Omid Ghamarnejad, Ragnar Bade, Paul Lukowicz, Jakob Karolus, Gregor Alexander Stavrou. "Beyond the visible: preliminary evaluation of the first wearable augmented reality assistance system for pancreatic surgery." *International Journal of Computer Assisted Radiology and Surgery*, 20(1), 117–129, 2025.

5. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, Jakob Karolus. "Design and Clinical Evaluation of ARAS: An Augmented Reality Assistance System for Open Pancreatic Surgery." In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 376–385. IEEE, 2024.

6. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, Jakob Karolus. "From Concept to Clinic: Multidisciplinary Design, Development, and Clinical Validation of Augmented Reality-Assisted Open Pancreatic Surgery." In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2025.

7. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, Jakob Karolus. "ARAS: LLM-Supported Augmented Reality Assistance System for Pancreatic Surgery." In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 176–180, 2024.

8. Javaheri, Hamraz, Omid Ghamarnejad, Rizky Widyaningsih, Ragnar Bade, Paul Lukowicz, Jakob Karolus, Gregor Alexander Stavrou. "Enhancing Perioperative Outcomes of Pancreatic Surgery with Wearable Augmented Reality Assistance System: A Matched-Pair Analysis." *Annals of Surgery Open*, 5(4): e516, 2024.

9. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, Jakob Karolus. "LLMs Enable Context-Aware Augmented Reality in Surgical Navigation." *DIS25: Designing Interactive Systems Conference*, 5–9 July 2025, Funchal, Madeira.

10. Javaheri, Hamraz, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, Jakob Karolus. "Breaking the Dogma: Misconceptions, Challenges, and Lessons Learned from AI and AR-Assisted Pancreatic Surgery." In *Proceedings of CHI '25 Workshop on Envisioning the Future of Interactive Health*, 2025.

Curriculum Vitae

Hamraz Javaheri

hamrazjavaheri@gmail.com
ORCID: 0000-0001-5494-0987

Education

2015–2017 **Electrical and Computer Engineering, Msc**
University of Kaiserslautern, Kaiserslautern, Germany

2011–2015 **Electronic and Communication Engineering, BSc**
Yidiz Technical University, Istanbul, Turkey

2013–2014 **European Union Student Exchange Program**
Aalto University, School of Electrical Engineering, Helsinki, Finland

Experience

2018– Present **Researcher** German Research Center for Artificial Intelligence
Embedded Intelligence, 67663 Kaiserslautern, Germany

2016–2017 **Research Assistant** German Research Center for Artificial Intelligence
Department of Embedded Intelligence, 67663 Kaiserslautern, Germany

2017 **Research Assistant** University of Kaiserslautern
Department of Educational Physics, 67663 Kaiserslautern, Germany

2017 **Research Assistant** German Research Center for Artificial Intelligence,
Department of Augmented Vision, 67663 Kaiserslautern, Germany

Academic Awards

Honorable Mention Paper Award	Designing Interactive Systems conference 2025, Funchal, Madeira
Best Demo Award	UbiComp / ISWC Conference 2024, Melbourne, Australia
Young Talent Award	Saarland Association of Surgeons 2025, Homburg, Germany

News Appearance

DW News	https://www.dw.com
ARD1	https://www.ardmediathek.de
Saarbrücker Zeitung	https://www.saarbruecker-zeitung.de