Breaking the Dogma: Misconceptions, Challenges, and Lessons Learned from AI and AR-Assisted Pancreatic Surgery

Hamraz Javaheri*
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Omid Ghamarnejad*
Allgemein-, Viszeral und Thoraxchirurgie, Chirurgische Onkologie, Klinikum Saarbrücken
Saarbrücken, Germany

Paul Lukowicz
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany

Gregor A. Stavrou†
Allgemein-, Viszeral und Thoraxchirurgie, Chirurgische Onkologie, Klinikum Saarbrücken
Saarbrücken, Germany

Jakob Karolus†
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany

Abstract

The growing body of research reflects the increasing interest and importance of integrating AI and AR into medical practice. Despite the rising prominence of these technologies in surgery, the proportion of publications in MEDLINE database (PubMed) has outpaced those in Human-Computer Interaction (HCI)-focused venues, with a noticeable decline in the HCI community's focus on critical medical applications. In this position paper, we (1) explore common misconceptions contributing to this decline, (2) draw insights from our recent projects and (3) share lessons learned from our experience in this evolving field.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; **Empirical studies in HCI**.

Keywords

Augmented Reality, Surgical Assistance System, Pancreatic Surgery, LLMs

ACM Reference Format:
Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor A. Stavrou, and Jakob Karolus. 2025. 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).* ACM, New York, NY, USA, 4 pages.

*Both authors contributed equally to this research.
†Both authors contributed equally to this research.

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan

1 Introduction

Over the past decade, significant advancements in artificial intelligence (AI) and augmented reality (AR) technologies have made a substantial impact across various sectors worldwide. Notably, in medicine, these innovations have the potential to significantly enhance patient outcomes, particularly in critical areas such as surgery [10, 13]. The growing body of research in this field reflects the increasing interest and importance of integrating AI and AR into surgical practice. A literature search spanning from 2010 to the present in the MEDLINE database (PubMed) reveals a striking trend: the volume of publications in the past three years (n = 34,483) has already surpassed the total publications from 2010 to 2021 (n = 33,312), underscoring the accelerating attention this topic is receiving within the medical research community (Figure 1).

Further analysis of various publication venues reveals an interesting shift in trends. Despite the growing prominence of these technologies in surgery, the proportion of publications in PubMed has outpaced those in Human-Computer Interaction (HCI)-focused venues, such as ACM Digital Library (DL) and IEEE Xplore (Figure 1). Specifically, 8,550 papers were published in the ACM DL from 2010 to 2021, whereas only 5,385 papers have been published from 2022 to 2025, suggesting a decline in the HCI community's focus on critical medical applications.

This observation raises the first question addressed in this paper: ***Why is there a decline in publications related to critical medical applications in the HCI domain?***

Furthermore, despite the rapid advancement of AI and AR technologies, many surgical methods still rely heavily on traditional techniques, some of which have remained unchanged for decades. This brings us to the second key question: ***Has the technology reached a level of readiness where it no longer requires further HCI research, or is there a disconnect in the adoption of these technologies in surgical practice?***

In exploring this question, we also aim to examine: ***Why are surgical procedures not advancing as rapidly as computing technologies, and who or what is responsible for this lag?*** In this paper, we aim to explore these critical questions, drawing insights

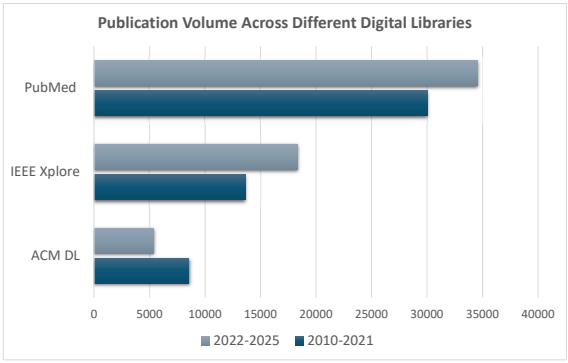

Figure 1: Publication Volume per Interval Across Different Digital Libraries

from our most recent project and offering lessons learned from our experience in this evolving field.

2 AI and AR Assisted Pancreatic Surgery

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 [3]. 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 [7, 11, 12]. 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 [1, 3].

Through multidisciplinary research involving research teams from computer science and visceral surgery, we addressed this research gap and developed ARAS, an AR assistance system for pancreatic surgery [9, 10]. After successfully implementing the system, we evaluated its usability and accuracy during clinical trials involving 20 patients with pancreatic tumors. Through a series of conducted studies, we demonstrated successful results, including the establishment of domain-specific design guidelines [9], the system's accuracy in registering patient-specific anatomical models [5, 7], improved perioperative outcomes compared to traditional surgical approaches [10], and the safe integration of large language models (LLMs) in surgical navigation systems [6, 8].

3 Misconceptions

In this section we will discuss our identified misconceptions, which we believe play a substantial role in decline in publications related to critical medical applications in the HCI domain.

3.1 Ethical Concerns: Who is the Judge?

One of the major concerns when conducting HCI-focused research in critical domains involving patient participation are ethics. Obtaining ethical approvals is a thorough and time-consuming process, especially for studies involving patients, as they often require approval from medical associations and regulatory bodies based on legal frameworks. This process is not feasible without the involvement of researchers from the medical field, who take responsibility for the study and ensure patient safety at all times.

Based on our experience, even with proper ethical approvals from both medical and institutional ethics boards, we have encountered concerns raised solely by reviewers from computer science community. These reviewers often advocated for paper rejection on the grounds of unethical research, despite the study adhering to all necessary ethical guidelines.

The major ethical concern raised was the issue of responsibility in cases where the system provides misinformation during clinical trials. However, what is often overlooked is surgeons' intellect and the fact that surgeons undergo years of specialized training and perform numerous procedures before developing the skills needed for complex surgeries without relying on navigational systems. Their role as system evaluators in the context of research does not interfere with their primary duty as surgeons, and it is necessary to further improve technological advancements.

The conundrum of insufficient prior evaluation being deemed unethical, while simultaneously avoiding further evaluation due to ethical concerns, creates a significant barrier to progress in these domains. With the rapid evolution of generative AI applications, it is crucial to test these approaches within a safe framework to assess their feasibility in the medical domain. However, misconceptions and barriers—often arising from a lack of understanding of multidisciplinary research and underestimating the role of medical experts in the process—continue to hinder further studies and real-world evaluations, ultimately delaying the integration of such technologies.

3.2 Multidisciplinary Research: Co-author or Participant?

A fundamental cornerstone in achieving a readiness level that enables the widespread adoption of AR and AI technologies in surgery is demonstrating the usability and advantages of these systems, not only in laboratory settings but also in real-world clinical environments. However, this presents a dilemma: a surgeon is unlikely to use a system that has not been proven effective in actual surgery, while an engineer or computer scientist cannot develop a safe and reliable system for widespread use without sufficient real-world testing. A solution to this dilemma goes through performing multidisciplinary research. Unlike many other domains where the target group can either participate in the study or serve as co-authors to avoid conflicts of interest, surgical research challenges the efficiency of this approach. A surgeon must know the system (co-author role) to have confidence in the system and take responsibility for testing it safely within a surgical environment (participant role). Therefore, **in multidisciplinary research in critical medical domains, co-authors can—and should—also serve as participants until the system is ready to be evaluated by independent subjects.**

One of the key issues we encountered is the failure to acknowledge this reality. The continuous involvement of the same team of surgeons throughout the design, development, and evaluation stages is essential, especially for research work that has never been

done before. However, reporting the same individuals as both co-authors and participants across different study phases often raises concerns about validity of results. This skepticism stems from the fact that many reviewers lack firsthand insight into the unique constraints and requirements of conducting research in surgical environments.

3.3 Benchmark: A Missing Consensus?

A key reason for the rejection of papers on medical systems in HCI and engineering venues, as well as the barrier in widespread use of already developed system in clinical routine, is the lack of clear benchmarks. This absence has led to misconceptions, especially in surgical AR and AI applications, where no consensus has been established yet.

3.3.1 Accuracy or Outcome? Although there are no established standards or consensus, a common misconception about medically approved navigation systems is the need for millimetric accuracy. This belief, often emphasized by researchers from both disciplines, overshadows the importance of positive outcomes. Since there are no standardized methods for measuring 3D spatial registration errors in AR systems used in surgical settings, expecting a metric report on system accuracy is unrealistic. Furthermore, error tolerance margins vary significantly depending on the purpose, intent, and the user's reliance on the system to complete a task. For instance, high registration accuracy is crucial in procedures like tumor ablations, where surgeons depend on the system for precision. However, in applications like surgical navigation, the system might just be used to provide an insight and some errors might be tolerated as surgeons can recognize and compensate for them. Hence, in evaluating such applications **subjective expert evaluations can often provide more meaningful insights than pure metric reports, especially when there is no consensus on the acceptable error margins nor a standard way of measuring it.** The presence of these misconceptions potentially led to more papers being accepted in medical fields rather than HCI or engineering focused venus, where often the emphasis is placed on the technology's impact on patient outcomes rather than solely on system metric accuracy.

3.3.2 Quality or Quantity? Another common misconception is focusing on the minimum number of participants required for a study, emphasizing quantity over quality. Throughout our research, we have observed a tendency to overlook the expertise and experience necessary for participation in studies focused on highly specialized and critical domains such as pancreatic surgery. Despite several sources emphasizing the importance of expert involvement in HCI research and acknowledging that participant numbers may differ from studies targeting broader audiences [2, 4], this misconception persists. Many reviewers still assume that studies with a lower number of participants, regardless of their expertise and specialization, contribute little to the research community and are not worthy of publication. **Focusing on quantity over quality when it comes to participants' profile only serves to delay progress and hinder advancement in highly specialized and critical domains.**

3.4 Novelty: What is Considered Novel?

Another major barrier to publishing HCI-related articles focused on critical surgical domains is the misconception around the concept of "novelty" which often becomes an excuse for paper rejections. **While the definition of novelty is hard to objectively grasp, research is not strictly required to be entirely novel, but it can serve as a reconfirmation of previous findings**. For example, using AR to overlay patient anatomy during surgery may not seem novel to the computer science community, as AR has been previously used in other domains to visualize spatial relationships. However, applying this approach for the first time in pancreatic surgery is considered entirely novel in the medical field due to the well-known challenges associated with this procedure. Most of the time applying previously used approaches from other domains to the medical field is frequently deemed insufficiently novel for publication by computer science community, despite offering new insights. On the other hand, validating entirely new approaches—such as using LLMs—might be considered unethical by some reviewers, even with proper ethical approvals, as these methods have not been extensively tested for use in critical domains involving patient care. As we have repeatedly emphasized the importance of conducting real-life studies in surgical contexts, this barrier ultimately limits the number of studies in ecologically valid environment and their publications in the HCI and more generally in computer science field.

3.5 Generalizability: Does One Size Fit All?

A common misconception among HCI community is assuming that study findings should be easily applied to other domains. Reviewers often raise this concern, sometimes leading to paper rejections. But the real question is: when designing a system for such a high-stakes field, should the goal be a one-size-fits-all solution, or should it be tailored to meet the unique needs of that domain? **It's unrealistic to expect someone using a system for general tasks to behave the same way as someone making life-or-death decisions under pressure and stress.** This misunderstanding shows a lack of insight into the field.

4 Lessons Learned

4.1 Communication Barriers

In multidisciplinary projects, particularly in surgery, it is crucial to conduct initial research from both perspectives. From the surgeon's standpoint, the problem and potential solutions must be clearly defined, while from the informatics or engineering perspective, it is essential to assess the feasibility of these solutions. A significant challenge that often arises during the process, is the failure of both parties to understand each other's terminology. Many terms in both fields do not have a direct equivalent, leading to confusion, reduced motivation, and hindering project progress. An important aspects in finding feasible solution which helps to faster integration of new technologies in medical domain is to **develop a common knowledge background and expand understanding of the field.** In our project, we have repeatedly experienced the importance of basic anatomical knowledge in understanding the problems stated by surgeons and in designing appropriate systems to address these issues.

Furthermore, clear communication between the two parties is essential from the outset to **set realistic goals and ensure that the design and development of the system are seen as an evolving process**, requiring continuous revisions throughout the project. Communication challenges between disciplines can often lead to poor system design that fails to meet the surgeons' requirements, ultimately resulting in demotivating research outcomes (cf. [14]). This, in turn, widens the gap between state-of-the-art advancements in computer science and their application in the medical field.

4.2 Intertwining Engineering and HCI

Throughout our experiences, we have learned that conducting meaningful real-life HCI research in such a specialized and critical domain requires starting with a robust system. While this was often assumed to be be not the primary concern for the HCI community, it is essential. **In research areas where no existing products guarantee system performance, developing a reliable system becomes a fundamental cornerstone of HCI research.** To achieve this, we found it especially important for engineers and designers designing the system to observe its real-time use in the operating room. In complex surgeries, such as pancreatic procedures—which typically last around four hours—system issues often arise during the operation. If not addressed immediately, these problems may be forgotten by the surgeon after the procedure. In our study, we closely observed the entire process, and implemented a detailed note-taking system during surgery, and recorded relevant discussions among the surgeons regarding the system's use throughout the procedure. This approach, combined with follow-up interviews after each surgery, enabled us to make significant improvements to the system. By the last ten surgeries, the system had stabilized, and no significant issues were encountered.

5 Future Direction: Why Do We Need More HCI Perspective?

To successfully integrate technological advancements into the surgical domain and ensure it evolves alongside the fast-paced progress of computing technologies, we must challenge existing dogmas and adopt a forward-thinking approach. Despite the hurdles, we remain firm in our belief that publishing in HCI domains offers transformative potential that is often overlooked. A key factor slowing the integration of technology in surgical environments is the lack of comprehensive, multidisciplinary HCI research in this area. As a result, systems are frequently poorly designed, failing to meet the demands of clinical routines and leaving both developers and the medical community frustrated. Now, more than ever, it is essential to close the gaps between these two disciplines. By doing so, we can harness the full potential of technological innovations and drive meaningful improvements in patient outcomes, ultimately revolutionizing healthcare.

References

[1] Manuel Birlo, PJ Eddie Edwards, Matthew Clarkson, and Danail Stoyanov. 2022. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. *Medical Image Analysis* 77 (2022), 102361.

[2] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.

[3] Benish Fida, Fabrizio Cutolo, Gregorio di Franco, Mauro Ferrari, and Vincenzo Ferrari. 2018. Augmented reality in open surgery. *Updates in surgery* 70, 3 (2018), 389–400.

[4] Wonil Hwang and Gavriel Salvendy. 2010. Number of people required for usability evaluation: the 10±2 rule. *Commun. ACM* 53, 5 (2010), 130–133.

[5] Hamraz Javaheri, Omid Ghamarnejad, Ragnar Bade, Paul Lukowicz, Jakob Karolus, and Gregor Alexander Stavrou. 2024. Beyond the visible: preliminary evaluation of the first wearable augmented reality assistance system for pancreatic surgery. *International Journal of Computer Assisted Radiology and Surgery* (2024), 1–13.

[6] Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. 2024. 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*. 176–180.

[7] Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. 2024. 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)*. IEEE, 376–385.

[8] Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. 2024. LLMs Enable Context-Aware Augmented Reality in Surgical Navigation. *arXiv e-prints* (2024), arXiv–2412.

[9] Hamraz Javaheri, Omid Ghamarnejad, Paul Lukowicz, Gregor Alexander Stavrou, and Jakob Karolus. 2025. From Concept to Clinic: Multidisciplinary Design, Development, and Clinical Validation of Augmented Reality-Assisted Open Pancreatic Surgery. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM. doi:10.1145/3706598.3713458

[10] Hamraz Javaheri, Omid Ghamarnejad, Rizky Widyaningsih, Ragnar Bade, Paul Lukowicz, Jakob Karolus, and Gregor Alexander Stavrou. 2024. Enhancing Perioperative Outcomes of Pancreatic Surgery with Wearable Augmented Reality Assistance System: A Matched-Pair Analysis. *Annals of Surgery Open* 5, 4 (2024), e516.

[11] Tomoyoshi Okamoto, Shinji Onda, Jungo Yasuda, Katsuhiko Yanaga, Naoki Suzuki, and Asaki Hattori. 2015. Navigation surgery using an augmented reality for pancreatectomy. *Digestive surgery* 32, 2 (2015), 117–123.

[12] Karl J Oldhafer, Gregor A Stavrou, Guido Prause, Heinz-Otto Peitgen, Tim C Lueth, and Stefan Weber. 2009. How to operate a liver tumor you cannot see. *Langenbeck's archives of surgery* 394 (2009), 489–494.

[13] Niki Rashidian, Mohammed Abu Hilal, Isabella Frigerio, Martina Guerra, Sigrid Sterckx, Francesca Tozzi, Giulia Capelli, Daunia Verdi, Gaya Spolverato, Aiste Gulla, et al. 2024. Ethics and trustworthiness of artificial intelligence in Hepato-Pancreato-Biliary surgery: a snapshot of insights from the European-African Hepato-Pancreato-Biliary Association (E-AHPBA) survey. *HPB* (2024).

[14] Yu Saito, Maki Sugimoto, Satoru Imura, Yuji Morine, Tetsuya Ikemoto, Shuichi Iwahashi, Shinichiro Yamada, and Mitsuo Shimada. 2020. Intraoperative 3D hologram support with mixed reality techniques in liver surgery. *Annals of surgery* 271, 1 (2020), e4–e7.