

Review

Open Access



The role of artificial intelligence in abdominal wall surgery: recent progress and embracing uncertainty

Toby Collins, Guinther Saibro, Antonello Forgione, Alexandre Hostettler, Jacques Marescaux

Department of Surgical Data Science, IRCAD France, Strasbourg 67000, France.

Correspondence to: Dr. Toby Collins, Department of Surgical Data Science, IRCAD France, 1 Place de l'Hôpital, Strasbourg 67000, France; E-mail: toby.collins@ircad.fr

How to cite this article: Collins T, Saibro G, Forgione A, Hostettler A, Marescaux J. The role of artificial intelligence in abdominal wall surgery: recent progress and embracing uncertainty. *Art Int Surg.* 2025;5:254-69. <https://dx.doi.org/10.20517/ais.2024.92>

Received: 17 Oct 2024 **First Decision:** 28 Nov 2024 **Revised:** 7 Apr 2025 **Accepted:** 15 Apr 2025 **Published:** 27 May 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Artificial intelligence (AI) is profoundly impacting most, if not all, scientific and medical disciplines. In abdominal wall surgery (AWS), which includes common procedures such as hernia repair, abdominal wall reconstruction, and separation, AI models trained on surgical data have immense potential to enhance clinical practice and patient outcomes. The benefits include better procedure planning, standardization, interventional guidance, awareness of critical structures, complication prevention, quality assurance, and patient monitoring. Moreover, AI may significantly transform surgical education by enhancing training, skill assessment, and feedback mechanisms, leading to better-prepared surgeons. This review article highlights the latest developments in AI and AWS, focusing on key emerging applications and why embracing AI model prediction uncertainty is essential to translating these research efforts to clinical practice.

Keywords: Artificial intelligence, surgery, hernia, computer-assisted surgery, machine learning

INTRODUCTION

Abdominal wall surgery (AWS) involves repairing defects or weaknesses in the abdominal wall, such as hernias, which may occur from congenital conditions, surgical incisions, or trauma. The goal is to restore the abdominal wall's structural integrity while minimizing complications to achieve positive functional and aesthetic outcomes. Each year, over 20 million inguinal hernia repairs are performed globally, making it one



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



of the most common surgical procedures^[1]. The hernia repair devices and consumables market has increased from \$4.7 billion in 2019 to a projection of \$6.4 billion by 2027^[2]. This is driven by advanced prosthetics, increasing adoption of robotic surgery, and a higher global prevalence rate^[3].

Abdominal wall hernias may be naturally occurring or acquired. Naturally occurring hernias include inguinal (groin area, most common), femoral (lower groin, more common in women), umbilical (near the belly button, common in infants and pregnancy), and epigastric (midline upper abdomen). Acquired hernias include incisional hernias, which develop at previous surgical sites due to weak healing, and parastomal hernias^[4], which form around a stoma due to abdominal wall weakness. Other types include Spigelian hernias (along the lateral abdominal muscles) and diastasis recti, a separation of abdominal muscles rather than a true hernia.

Hernia surgery often requires specialized skills beyond the scope of a general surgeon^[5], and the challenges are attributed to several factors:

Complex anatomy: The groin region is considered anatomically complex^[6], and precise anatomical knowledge is critical to avoid injury to nerves, blood vessels, or organs.

Mesh complications: Prosthetic meshes are frequently used in hernia repair; however, they introduce risks of infection, rejection, or adhesion to underlying organs. Surgeons must choose the appropriate mesh type, size, and placement method. The mesh must be suitably fixed, and the use of staplers or tackers, if placed incorrectly, carries greater risks of chronic postoperative pain^[7-9].

Large or complex defects: Repairing large or recurrent hernias or abdominal wall defects requires advanced techniques, such as component separation, to provide adequate and durable prosthetic support to prevent recurrence.

Tissue tension and closure: Achieving a tension-free closure is vital to prevent recurrence^[10], yet this can be difficult, especially in larger defects or after multiple surgeries, as tissue may be compromised or retracted.

Obesity and comorbidities: Patients with obesity or other conditions such as diabetes pose additional challenges^[11] due to thicker abdominal walls, a higher risk of infection, and difficulty in achieving a tension-free repair.

To address the above challenges, surgeons must carefully balance technical precision, material selection, and personalized patient care to improve outcomes. Specialized training, such as WebSurg's Hernia Basecamp^[12], might be an important factor in reducing recurrence and complication rates, in addition to advances in material science in mesh design, biomaterials that support regenerative medicine, and immune engineering^[13]. However, despite these initiatives and research directions, long-term follow-up of hernia patients is often disappointing, both in terms of complications and reoperation rates that remain unacceptably high (8% to 15%^[1,14,15]).

AI technologies have greater potential to improve abdominal wall and hernia surgery^[16,17], including applications in preoperative planning and risk assessment, intraoperative guidance and safety enhancement, postoperative monitoring, and surgery training. However, AI adoption in routine clinical practice has been relatively slow compared to generative AI models, such as ChatGPT, in broader society. As of today, while research in AI and AWS has advanced, there are not yet certified AI devices specifically designed to assist

with AWS. The following factors contribute to the absence of clinically translated solutions:

Regulatory challenges: One of the primary obstacles preventing the certification of AI devices for AWS is the rigorous regulatory landscape in healthcare. Regulatory agencies such as the FDA and EMA require that AI-based medical devices undergo extensive validation and testing before they can be approved for clinical use. The approval process for medical AI technologies is complex, often involving long timelines and substantial evidence to demonstrate safety, efficacy, and clinical relevance. Given the complexity of AWS and the variability in patient presentations, demonstrating that AI systems can reliably function across all potential scenarios is a significant challenge.

Complexity of surgical data: Unlike fields like radiology, where imaging data follow more standardized protocols, surgical data - especially video data from abdominal wall surgeries, are highly complex and variable. This variability arises from variations in camera angles, visibility issues due to bleeding or smoke, and differences in surgical techniques, instruments, and environments. These factors make it difficult to develop robust AI models that can reliably interpret surgical situations in real time. Moreover, the lack of standardization in surgical data limits the ability of AI systems to generalize across different hospitals, teams, and surgical settings.

Risk and accountability: Surgery is a high-stakes environment where mistakes can have serious consequences. For an AI system to be certified for clinical use in AWS, it must demonstrate a high level of accuracy, reliability, and decision-making support without compromising patient safety. The important question of accountability arises - if an AI system were to make an incorrect recommendation, determining responsibility for the resulting harm is complex.

Integration into clinical workflow: The integration of AI into real-world surgical environments presents significant logistical challenges. Most hospital systems are not optimized to incorporate AI tools seamlessly into day-to-day practice. AI systems may require integration with existing electronic health record (EHR) systems and operating room (OR) technologies. Additionally, surgical teams must be trained to trust and responsibly operate AI tools, which requires a significant investment in time and resources, and performance monitoring systems must be in place to ensure the AI system continues to perform as intended. The process of integrating AI into a hospital's workflow, ensuring smooth communication between AI systems and surgical teams, and aligning these tools with hospital protocols is a substantial barrier to clinical translation.

Our article focuses on state-of-the-art research and emerging AI technologies that are currently being researched to assist AWS, rather than their immediate clinical implementation. While AI is not yet in routine clinical use for hernia repair, there is promising research underway that could eventually improve surgical outcomes.

ARTIFICIAL INTELLIGENCE BACKGROUND

Key concepts

Artificial intelligence (AI), machine learning (ML), and deep learning (DL) are interconnected fields that have revolutionized various industries and research areas in recent years. AI encompasses the broad concept of machines performing tasks that typically require human intelligence, such as problem-solving, language comprehension, and decision making. ML, a subset of AI, focuses on developing algorithms that enable machines to learn from data and improve over time without explicit programming. DL^[18], in turn, is a specialized and highly successful form of ML that utilizes computational models, known as artificial neural

networks (ANNs), which are broadly inspired by the structure and function of the human brain.

ANNs consist of artificial neurons arranged in layers. Each neuron has input connections to receive data and output connections to transmit processed information to subsequent neurons. Every connection between neurons has a weight, which determines the strength of the information passed into the neuron. These weights are crucial to the network's performance, and during training, the weights are automatically adjusted to maximize the model's performance according to a performance metric, which is selected according to the task at hand, such as anatomical structure recognition accuracy.

A single artificial neuron processes its inputs according to simple mathematics; in general, the inputs are multiplied by their weights, then the weighted inputs are added together, and then an activation function is applied, which determines whether the neuron should activate or remain inactive based on the input it receives. The activation function plays a crucial role in allowing neural networks to learn and make decisions by introducing non-linearity to the output. Various activation functions exist. For example, the sigmoid activation function maps the input to a value between 0 and 1, acting like a smooth "on/off" switch that controls the strength of a neuron's output. This makes it useful for tasks like binary classification, where the output can represent probabilities.

While a single artificial neuron has limited practical use, the complexity and capability of an ANN arise by using many interconnections between many artificial neurons. In particular, increasing the depth of the network (i.e., the number of layers of neurons) allows ANNs to learn complex patterns from training data, enabling tasks such as medical image analysis^[19], language translation^[20], and generative AI applications^[21,22] like image, video^[23], and speech^[24] creation. The size of contemporary ANNs varies significantly, primarily depending on the range of tasks they are designed to perform - broader task ranges typically require larger ANNs. The largest ANNs used for understanding and generating natural language are large language models (LLMs), such as OpenAI's GPT-4 (used in ChatGPT^[25]). While OpenAI has not released official statistics, some open-source alternative LLMs - such as Meta's Llama 405B - report having 405 billion trainable parameters. In contrast, AI models intended for use as medical devices tend to be smaller due to narrower intended use. For instance, models used to delineate structures in radiology images, such as nnUNet^[26], typically have tens of millions of trainable parameters, requiring substantially less time, computational resources, and financial investment for model training.

The rapid growth of AI, ML, and DL in recent years is attributed to several key factors. First, the explosion of data in the "big data" era has provided essential training material for advanced DL models. Second, significant advancements in computational power - fueled by powerful GPUs and specialized hardware - have enabled faster processing and more efficient model training. Third, the emergence of open-source, sophisticated programming frameworks like TensorFlow^[27] and PyTorch^[28] has made it easier for researchers and AI engineers to train complex models. Lastly, increased investment in AI research and development, alongside growing interest from both the public and private sectors, has accelerated progress in these fields.

AI model development for healthcare applications

AI development for healthcare applications typically progresses through five main phases. The first phase, data collection and model training, involves gathering high-quality, relevant data to teach the AI model. A critical aspect of this phase is data annotation, where input-output pairs are labeled or "tagged" with specific information, enabling the model to learn in a process known as supervised learning. This process is similar to how a student studies for an exam by practicing with questions, looking at the correct answers, and then

learning from their mistakes. During supervised learning, the initial “untrained” AI model makes predictions on the training data, and a performance metric is used to evaluate how well it performs, such as sensitivity, specificity, precision, recall, accuracy, or F1 (the harmonic mean of precision and recall). Based on this evaluation, an automatic training process is launched, which makes small modifications to the model’s parameters to improve its performance metric. The process repeats iteratively, and the model gradually improves its ability to make accurate predictions on the training data.

Alternative ML methods exist, and they can be used alongside supervised learning to enhance model training. One such approach is self-supervised learning^[29], where the model generates its own labels from the input data, allowing the model to identify inherent structures or relationships in the data. Self-supervised learning is often used as a prelude to supervised learning, where its main benefit is to reduce the amount of required labeled data. Another notable approach is reinforcement learning (RL), which complements supervised learning by enabling the model to learn through trial and error. In RL, the AI model learns strategies, or decision-making policies, designed to maximize a virtual reward. A virtual reward refers to a feedback signal that the model uses to gauge the effectiveness of its actions. Unlike in supervised learning, where the model is provided with labeled training data, in RL, the model explores actions and receives rewards (or penalties) based on the success or failure of those actions. This virtual reward can be a numerical value that increases when the model takes actions that bring it closer to achieving its goal, or decreases when it moves further away from the desired outcome. Over time, the model learns to maximize these rewards by refining its action/decision-making policy.

To draw an analogy, RL is similar to training in a surgical simulation. Initially, the surgeon might perform steps in a procedure without knowing whether the action is optimal or whether the outcome will be successful. However, the simulator provides feedback that may be positive (if the surgeon performs a task correctly, reinforcing the trainee’s good decisions/actions) or negative (e.g., if an action leads to a complication), prompting the surgeon to adjust their approach on their next trial. As the surgeon continues to practice, they refine their skills based on this continuous cycle of action, feedback, and adjustment, ultimately leading to mastery. Similarly, in RL, the model continuously learns from its experiences by receiving virtual rewards, allowing it to optimize its performance over time. The main advantage of RL over supervised learning is that RL allows the model to learn optimal decision-making strategies through trial and error, without requiring explicit labels to tell it which actions or decisions should be performed at each step.

The second phase of AI model development is internal validation^[28], where the performance of a trained AI model is tested using data similar to the training data, typically collected concurrently with the training dataset. This helps catch common problems early, such as overfitting or model bias. Overfitting happens when the model learns random details or noise in the training data instead of general patterns, causing it to perform well on the training data but poorly on new, unseen data. Model bias, on the other hand, occurs when the model makes incorrect assumptions about the data, leading to inaccurate predictions.

The third phase is external validation, where the model is tested on unseen, independent data. This step is crucial for assessing the model’s ability to generalize to new, real-world scenarios and ensuring that it can perform reliably outside the training environment. The fourth phase involves the regulatory approval process, where models must meet strict standards and regulations regarding performance, intended population, and safety before they can be deployed for real-world use^[30]. Finally, the fifth phase is performance monitoring, which focuses on tracking the model’s performance over time. This phase is required to detect “drift”, or any decline in performance caused by changes in the data or the environment

that the model was not originally trained on. Monitoring also helps identify issues like incorrect data labeling or shifts in regulatory requirements, prompting updates to the model or its application to ensure ongoing reliability and compliance.

Although we have outlined the primary steps involved in AI model development, a detailed discussion falls beyond the scope of this article. For a comprehensive examination of AI model development in healthcare, we refer readers to a selection of foundational texts, including^[31,32].

RESEARCH IN AI AND AWS

Researchers from various other institutions are advancing AI technologies specifically to enhance AWS, requiring close interdisciplinary collaborations among surgeons, AI researchers, software engineers, and clinical researchers. Drawing on our own experience, this article explores how advancements in AI technologies can enhance AWS, organized by five key application areas: (1) preoperative planning and outcome prediction; (2) intraoperative assistance; (3) emergency surgery; (4) postoperative care; and (5) surgery education. This section reviews the literature, divided into the above five categories.

Preoperative planning and outcome prediction

Computed tomography (CT) is generally considered the most effective imaging modality for accurately assessing the size and location of abdominal hernias, rectus diastasis, and any associated muscle atrophy, as well as the hernia's proportion relative to the abdominal wall. This detailed information can guide surgeons in selecting the most appropriate surgical approach [open vs. minimally invasive surgery (MIS)], determining the optimal positioning and fixation of meshes, and assessing the need for additional interventions, such as botulinum toxin injections, preoperative pneumoperitoneum, or component separation techniques.

However, despite their importance, these findings are often missing in CT reports, as radiologists may not be familiar with evaluating the abdominal wall or understanding the specific details required for surgical planning. For this reason, surgeons often do not consult CT images, relying instead on physical examination and patient history.

The underutilization of preoperative CT has been recognized as a limitation of current practice, driving recent efforts to improve radiology hernia reports for surgeons^[33]. AI has great potential to accelerate this process. An important advance is the concept of a 3D virtual clone (or “digital twin”)^[34], where AI models (typically deep neural networks - DNNs^[35]) reconstruct 3D models of a patient's specific anatomy from a preoperative image, with a process known as image segmentation. Neural networks such as nnUNet^[26] can delineate a wide variety of tissues, and recently, TotalSegmentator^[36], adapted from nnUNet, delineates 117 anatomical and pathological classes in CT images using a training dataset of 1,228 patients. Digital twin research has matured with products approved for medical use. Visible Patient, a startup founded in 2013 that spun off from IRCAD France, provides an online service to create patient-specific anatomy models^[34] from CT and MR images. This combines DNNs and human oversight for output correction and validation.

There are several preliminary works to generate automatic patient-specific 3D models of hernias^[37,38]. Zhang *et al.* developed a method to differentiate abdominal wall tissue from the hernia sac to improve 3D hernia visualization and measurement and assist treatment planning^[37]. This was a small pilot study; however, mature models for finer-grained differentiation of muscles, fascia, hernia sacs, and vessels will soon exist, trained on larger cohorts with robust multi-centric external validation.

The use of preoperative imaging and digital twins extends toward other applications in AWS, including surgical approach complexity prediction^[39] and recommendation, as well as prediction of postoperative complications such as surgical site infection (SSI) and pulmonary failure^[39,40]. The “surgical complexity” prediction performance in^[39] was encouraging and reportedly better than expert surgeon judgments, with an accuracy of 81.3% compared with 65.0% ($P < 0.001$). However, the validation of this study has several limitations that impact the generalizability and reliability of its findings. One key concern is the lack of external validation, as the models were only tested within the same institution where they were developed. Without applying the deep learning models (DLMs) to independent datasets from different hospitals or more diverse patient populations, it is unclear whether they would perform as well in broader clinical settings. Additionally, the patient cohort was mostly white (87.5%), which limits the applicability of the models to more diverse populations. The small validation set for the surgical complexity model (35 patients) further raises concerns about the robustness of the reported accuracy and predictive power. While internal validation is useful, a larger and more heterogeneous dataset would provide stronger evidence of the model’s reliability. Another notable limitation is the lack of expert comparison for the SSI and pulmonary complication models. Unlike the surgical complexity model, which was evaluated against expert surgeon predictions, these other models were not benchmarked against clinical judgment, making it difficult to assess their real-world utility. Furthermore, the pulmonary complication model performed poorly, with the area under the receiver operator curve (ROC-AUC) of only 0.545, suggesting that the features extracted by the model may not be clinically meaningful. This raises questions about the adequacy of the training process and feature selection.

Outcome prediction in healthcare, particularly in complex cases like mesh infection, presents a significant challenge due to the multifaceted nature of the factors involved. One major difficulty is the need for large-scale, longitudinal datasets, as outcome data such as mesh infection may not manifest until months or even years after the initial procedure^[41]. These delayed outcomes make it crucial to gather and analyze data over extended periods, often involving patients’ health trajectories, lifestyle changes, and treatment responses over time.

Moreover, effectively predicting these outcomes requires the integration of diverse data sources. For instance, preoperative imaging can provide insight into the patient’s anatomy and potential risks associated with surgery, while EHRs offer a comprehensive history of the patient’s medical conditions, previous treatments, and comorbidities. Blood biomarkers, genetic information, and physiological data from wearable devices can reveal critical details about the patient’s internal processes and response to treatment, contributing to a more personalized and accurate prediction model.

In addition to these external data sources, intraoperative factors - such as the surgeon’s skill, experience, and technique - play a vital role in the success or failure of a procedure, further complicating the prediction process. The combination of all these data sources can provide a holistic view of the patient’s condition, but it also introduces challenges in data integration, quality control, and analysis. More research into advanced AI and ML models, which can handle such complex, multidimensional datasets, is essential for improving the accuracy of outcome predictions, ultimately helping healthcare professionals make better-informed decisions.

Intraoperative assistance

AI may contribute to intraoperative assistance in AWS in several ways:

Enhanced surgery visualization and safety: AI systems can integrate data from multiple sources to provide surgeons with enhanced visualization. For instance, we have developed AI models to automatically detect and track seven key anatomical structures in transabdominal preperitoneal (TAPP) inguinal hernia repair^[16]. This AI system, which can integrate with robot arm movement data, may be used to continually visualize structures even if they are obscured by bleeding or a prosthetic. This visualization could enhance safety in inguinal hernia surgery by reducing complications, especially chronic postoperative pain due to improperly located prosthetic staples, and preventing vascular injury near the “Triangle of Doom”.

Procedure monitoring, standardization, and quality assurance: Procedure guidelines, such as the “10 Golden Rules” in MIS inguinal hernia repair^[42], were proposed to improve hernia repair quality and standardization. However, they are not widely adopted due to a lack of awareness/training, clinical practice variability, and guideline complexity. AI systems may significantly increase guideline adoption^[16] by continually monitoring surgical video data and validating guideline conformity in real time.

We have made the first advances in this direction by developing an AI system that detects critical structures that must be exposed to ensure a complete TAPP inguinal hernia dissection^[16]. This model may be integrated with a mesh detection model to alert the surgeon if they attempt to place a mesh over an inadequately dissected hernia. We have also developed an AI model to divide TAPP surgical videos into seven main surgical steps automatically^[17]. The steps may be nonsequential, typical in bilateral repairs, and the system can process live and archive videos. It has various applications, including quantifying time spent on each step, which can be applied to enhance OR efficiency by, e.g., alerting the support staff to prepare the mesh when the dissection is nearing completion. In the future, these AI models will be enhanced in analysis depth, e.g., quantitative assessment of dissection zones, mesh coverage, and tissue tension, and their application breadth for use in different abdominal wall procedures.

Automated surgery reports: AI systems are being developed to analyze surgical images and videos and to generate automated reports^[43]. They are likely to be adapted to AWS soon. This may reduce the burden on the surgical staff while also providing a comprehensive and systematic surgical record of every procedure. Using video-based recognition models, they may also provide information beyond that of standard reports, including a detailed log of events and surgical actions/decisions.

Emergency surgery

Abdominal wall emergencies, such as incarcerated or strangulated hernias, pose significant challenges that differ from elective procedures. In these situations, time is critical, and surgeons must make rapid decisions to prevent complications like bowel ischemia, perforation, and sepsis. Unlike planned surgeries, where there is time for thorough preoperative imaging and detailed surgical planning, emergency cases often require immediate action based on clinical judgment and limited diagnostic information.

One of the most pressing concerns in emergency AWS is determining whether bowel loops trapped within a hernia sac remain viable or need resection. This decision is traditionally made based on visual inspection and subjective assessment of factors like bowel color, peristalsis, and pulsation. However, misjudging bowel viability can lead to either unnecessary resections or, conversely, leaving necrotic tissue in place, both of which can result in severe complications. AI-assisted imaging techniques, such as real-time perfusion analysis using near-infrared fluorescence^[44,45] or hyperspectral imaging^[46,47], could help surgeons make more accurate assessments and reduce errors in intraoperative decision making.

Another major challenge in emergency settings is deciding whether to use prosthetic mesh for hernia repair, particularly in contaminated fields where the risk of infection is high. In elective procedures, careful patient selection and preoperative optimization help mitigate these risks, but in emergencies, there is often little time for such preparation. AI-driven predictive models could help guide decision making by analyzing patient-specific risk factors, surgical conditions, and historical outcomes to suggest the safest and most effective repair approach.

Patient fragility is another key factor influencing surgical outcomes in emergencies. Many patients presenting with abdominal wall emergencies are elderly or have significant comorbidities such as diabetes, cardiovascular disease, or immunosuppression, making them more vulnerable to complications. AI-based risk stratification tools could assist surgeons in determining the best course of action, balancing the risks of immediate surgery against the potential for conservative management in high-risk patients.

Postoperative complications are more common in emergency AWS than in elective cases. Issues such as SSIs, wound dehiscence, and anastomotic leaks can significantly prolong hospital stays and increase morbidity. AI-driven early warning systems, which continuously monitor vital signs, laboratory results, and other patient data, could help detect complications at an earlier stage, allowing for timely intervention and potentially improving patient outcomes.

Postoperative care

The postoperative phase is crucial for ensuring smooth recovery and early detection of complications following AWS, particularly in hernia repair and reconstruction. There has been limited research in this area specific to AWS; however, we expect systems to be developed focusing on the following key areas:

Vital Sign Monitoring: AI algorithms can analyze real-time data from wearable sensors to track key physiological parameters, such as heart rate, oxygen saturation, and body temperature^[48]. In the context of AWS, they could be valuable tools for early detection of complications like infection or respiratory distress.

Wound Assessment: Postoperative wound care is critical in hernia repair to prevent complications such as infection or poor healing. AI-driven image recognition tools can be used to monitor the surgical site, detecting early signs of infection, dehiscence (wound opening), or excessive swelling^[49-51]. For instance, AI can analyze visual data from wound images to identify abnormal signs like increased redness, warmth, or fluid accumulation, allowing clinicians to intervene early and reduce the risk of wound-related complications.

Mesh Complication Detection: Mesh complications in hernia repairs can include abscesses, hematomas, seromas, fistulas, bowel obstructions, mesh retraction, granulomas, and recurrent hernias. Imaging techniques, particularly ultrasound^[52] and CT^[53], are valuable tools for diagnosing these complications at earlier stages, allowing for timely intervention and adjustment of treatment plans. Over the past few years, the integration of AI into medical imaging has significantly transformed healthcare^[54]. We believe these technologies should be incorporated into routine postoperative examinations for mesh complication detection. By doing so, AI can enhance detection rates, reduce the burden on expert radiologists, and improve overall efficiency in patient care.

Personalized Recovery Plans and Patient Follow-up: By leveraging AI for personalized recovery planning, postoperative care for abdominal wall reconstruction and hernia repair patients can be more efficient and tailored to individual recovery processes. AI-driven models analyze factors such as preoperative health,

surgical complexity, lifestyle, genetics, and real-time recovery data to create customized recovery plans. These plans could:

- Optimize recovery trajectories: Where AI systems propose adapting the recovery plan in real time, ensuring pain management^[55], physical therapy, and wound care are tailored to the patient's progress.
- Reduce hospital stays: Where recovery metrics are monitored in real time, leading to AI systems helping recommend when patients are ready for discharge^[56], reducing unnecessary hospital stays while maintaining optimal recovery outcomes.
- Offer customized exercise recommendations: Where AI systems could make personalized exercise recommendations to strengthen abdominal muscles while ensuring the integrity of the surgical repair.

Virtual health assistants: AI is also being explored to improve patient engagement and postoperative follow-up through automated systems and virtual health assistants. LLMs integrated into conversational AI systems are being investigated to facilitate routine follow-ups^[57], answer common patient concerns, and guide them through the recovery process. AI-powered rehabilitation programs can monitor patient mobility, guide core-strengthening exercises, and provide real-time feedback to prevent strain on the healing abdominal wall. These interventions can be particularly beneficial for high-risk hernia repair patients who require specialized recovery strategies.

Surgery education

The roles of AI in surgery education are likely to be wide-ranging, from realistic simulation-based training^[58,59], creation of personalized learning pathways, and video-based competency assessment^[60] to AI-curated case studies, technical skill assessment^[17], and the enhancement of video-based learning platforms such as WebSurg^[61]. Additionally, LLMs such as ChatGPT can process, generate, and synthesize vast amounts of medical knowledge, making them increasingly valuable tools in surgery education. A very promising future direction is the integration of LLMs with surgery video education to enable students to learn about video content using natural conversational language. To this end, there has been some recent research in Visual Question Answering, such as Surgical-VQA^[62], which is an AI model trained to answer questionnaires on surgical procedures based on the procedure's video.

The use of LLMs regarding surgery, however, extends far beyond education, as indicated in the previous section. The applications range from improving patient communication^[63], helping document clinical research^[64], performing statistical analysis, and identifying study biases^[65] to generating code to assist the development of medical device software^[66].

DISCUSSION

From research to clinical deployment

Advancing from research to clinical deployment of AI in AWS

The application of AI technologies to AWS is in its infancy. As highlighted above, numerous applications are being investigated in the research literature; however, none have yet translated into routine practice.

In the development of AI technologies for AWS and hernia repair, including our own research^[16,17], pre-recorded surgical videos have been instrumental in model development and validation. While these retrospective datasets provide a strong foundation, they also present key limitations.

One major challenge is that AI models are typically evaluated based on task performance metrics, such as structure recognition rates, without direct assessment of their real-world clinical impact. Transitioning from retrospective validation to real-world implementation in AWS is a complex, multi-step process.

AI models must be seamlessly integrated into the clinical workflow to provide real-time assistance without disrupting surgical autonomy or established protocols. A critical step in this transition is prospective clinical validation, often through randomized controlled trials (RCTs) or cohort studies. These studies would compare AI-assisted hernia repairs to traditional methods, assessing key outcome metrics such as complication rates (e.g., infection or recurrence), surgical efficiency (e.g., operation time), and long-term patient recovery, including pain management and return to daily activities. However, conducting such studies is time-intensive and resource-demanding. Moreover, integrating AI into AWS requires compatibility with hospital IT infrastructure, including EHRs and surgical planning tools.

It is also important to recognize the importance of developing AI systems that are robust enough to adapt to anatomical variations. One strategy to improve AI's adaptability in these cases is the inclusion of diverse datasets during the training phase, incorporating a wide range of anatomical variations and patient-specific characteristics. While this mitigates the issue, it may not completely eliminate it. Continuous monitoring and recalibration of AI models, based on real-time feedback and postoperative data, will help improve the model's ability to provide accurate guidance across diverse patient populations. A promising future research direction is to develop AI systems to flag cases where anatomical features deviate significantly from the norm, prompting the surgeon to make additional assessments or adjustments.

To ensure accountability, prospective validation studies must meticulously document surgical decisions both with and without AI assistance. This allows researchers to directly attribute any observed outcome changes to AI intervention. Furthermore, long-term follow-up is crucial, as AI's true impact may take time to manifest. In hernia repairs, for instance, complications like chronic pain, mesh-related issues, or recurrence may only emerge months or years post-surgery.

AI technologies in AWS, particularly hernia repair, hold great promise, but significant cost-effectiveness challenges remain. While university research often produces the initial R&D results and methods, these innovations are typically commercialized by medtech companies, which fund further development, clinical trials, and validation. Hospitals, while not responsible for the R&D costs, must still invest in integrating these technologies into their existing infrastructure. This includes upgrading hardware, software, and training staff - expenses that can be substantial, especially if immediate benefits do not translate into clear financial returns.

A critical barrier to AI adoption in hernia repair is insurance reimbursement. Insurers may hesitate to cover these higher-cost procedures without clear evidence of cost savings or improved outcomes. The long-term benefits of AI - such as reduced follow-up surgeries or complications - may not be immediately reflected in short-term cost metrics, making it harder for hospitals to justify the additional investment. Reimbursement models need to adapt to consider both short-term expenses and long-term benefits, aligning the financial interests of hospitals, insurers, and medtech companies. Hospitals require clear financial incentives, including reimbursement models that reflect both immediate costs and long-term benefits. Demonstrating AI's ability to improve patient outcomes and reduce long-term healthcare costs will be essential in overcoming cost-effectiveness and reimbursement challenges, ensuring the broader adoption of AI in AWS.

Leveraging AI models with managed uncertainty: imperfection does not invalidate utility

It is important to state that in general, an AI model does not need to be perfect to be useful. Any ChatGPT user will attest to that. Indeed, if this were not true, research in AI and surgery should have stopped long ago, because the complexity of surgery data and surgical tasks practically ensures that AI model outputs will have at least some uncertainty. The AI systems described in this article will be deployed in clinical practice

only if the risks associated with uncertain predictions have been established in clinical trials and the benefits to the patient outweigh the risks. Understanding, communicating, and managing the risks associated with prediction uncertainty will be key to success.

AI model prediction uncertainty arises from two main factors: epistemic uncertainty and aleatory uncertainty. It is crucial for all stakeholders and users of AI, including surgeons making critical decisions based on AI results, to understand these concepts.

Epistemic Uncertainty: This type of uncertainty comes from a lack of knowledge or information. It is considered reducible, meaning that adding more training data can help the model better capture underlying patterns or relationships, leading to a more accurate representation of the system.

Aleatory Uncertainty: In contrast, aleatory uncertainty is caused by the inherent randomness and variability present in the data. This type of uncertainty is irreducible - no matter how many additional data are collected, the inherent variability in the data remains.

Considering epistemic uncertainty, if an AI system is trained on a dataset that lacks diversity in terms of patient demographics or surgical techniques, or patient abnormalities due to prior surgeries, surgery technique recommendations may be untrustworthy due to high epistemic uncertainty. Ensuring that AI models are trained on diverse and representative datasets reduces epistemic uncertainty, and the only way to understand if the uncertainty is sufficiently reduced is through rigorous model testing with appropriate external validation^[67].

However, reducing epistemic uncertainty alone is not sufficient to guarantee trustworthy results. For instance, a digital twin's accuracy is limited by the imaging modality from which it was generated. Various anatomical structures may not have sufficient contrast for unambiguous delineation, especially in non-contrast-enhanced CT. Consequently, a digital twin may be imperfect, no matter how many data are used for training. This is an example of aleatory uncertainty. By definition, this is irreducible, and whether it is a problem depends, crucially, on the intended use of the model's predictions. Before investing considerable research effort in developing a model and gathering training data, this should be fully understood.

Pragmatically, one should always estimate the aleatory uncertainty before investing time in AI model development. One of the best ways to do this is to have multiple expert surgeons perform the task independently and then measure their inter-rater agreement. A high disagreement indicates high aleatory uncertainty, which may prohibit the possibility of a useful AI model ever being developed. We performed this kind of analysis when we were researching our system to automatically verify the completion of the critical view of the myopectineal orifice (CVMPO)^[16], which helped steer the research toward a more focused goal of detecting critical structures in the CVMPO. Initially, we aimed to automatically verify the completion of every step according to CVMPO guidelines^[68]. However, our preliminary inter-rater analysis revealed that several steps, such as assuring dissection clearance of 2 cm from Cooper's ligament, had a low agreement among surgeons. This was due to the high difficulty in visually determining the absolute size of structures in a scene from single-lens (monocular) laparoscopic images. Instead, we focused on detecting critical structures, including Cooper's ligament, which was a substantially more objective task (with less aleatory uncertainty). Aleatory uncertainty can sometimes be reduced by altering the nature of the data - for example, by using contrast-enhanced CT scans to improve image clarity and image segmentation performance. However, in other situations, reducing this uncertainty may not be feasible without disrupting the surgeon's workflow or making the process impractical. For instance, we have shown that AI models can

recognize nerves below the surface using data from a hyperspectral camera^[69], and more accurately compared to a standard color camera used in laparoscopy. This has the potential to improve the safety of inguinal hernia repair. However, this camera is a specialized device that is more expensive and bulkier than a standard camera used in MIS, and it does not provide high frame rates. So, on the one hand, aleatory uncertainty can be reduced using a specialized camera, but on the other hand, one must balance this with any associated impracticalities or costs.

While various AI prediction models have been proposed in the literature, the next phase, before clinical deployment, should also involve developing models that can convey accurate assessments of their prediction uncertainty using model calibration^[70]. Calibration is required to ensure high uncertainty is communicated when a prediction is likely to be incorrect.

In this sense, the AI models serve as a tool to assist surgeons, rather than as a replacement for human expertise, who will need to assimilate the predictions and uncertainty information in their decision-making processes. AI models will continue to improve, yet adequate training of surgeons to use them as tools, with a transparent understanding of their inherent limitations, will be fundamental to success. This will require a collaborative effort on both sides. AI models will be required to provide results and justification for their results - known as “explainable AI”^[71]. AI models should also convey the sources of uncertainty.

On the other hand, surgeons will need additional training in data science and ML so they may understand how to interpret AI-generated outputs and safely integrate AI into their decision-making processes. This is not a simple feat. In radiology, recent studies show that, despite training, incorrect predictions can negatively influence a clinical decision due to over-reliance on the technology. For instance, radiologists were more likely to miss pathological findings in mammograms when an AI system also missed them^[72].

As AI technology advances, its application in AWS is expected to increase, contributing to more personalized and efficient surgical practices. Surgeons must remain informed about these developments and adapt to emerging technologies to optimize patient outcomes while upholding established standards of care.

DECLARATIONS

Authors' contributions

Initial manuscript drafting: Collins T

Manuscript revisions, proofreading and finalization: Saibro G, Forgione A, Hostettler A, Marescaux J

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Collins T is a Junior Editorial Board member of *Artificial Intelligence Surgery*. Marescaux J is the Honorary Regional Editor of *Artificial Intelligence Surgery*. Collins T and Marescaux J are Guest Editors of the Special Issue, *The Role of Artificial Intelligence in Abdominal Wall Surgery*. Collins T and Marescaux J were not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, and decision making. The other authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. HerniaSurge Group. International guidelines for groin hernia management. *Hernia*. 2018;22:1-165. DOI PubMed PMC
2. Allied Market Research. Hernia repair devices and consumables market size, share, competitive landscape and trend analysis report, by product, surgery type and hernia type : global opportunity analysis and industry forecast, 2020-2027. Available from: <https://www.alliedmarketresearch.com/hernia-repair-devices-market>. [Last accessed on 17 Apr 2025].
3. Wang F, Ma B, Ma Q, Liu X. Global, regional, and national burden of inguinal, femoral, and abdominal hernias: a systematic analysis of prevalence, incidence, deaths, and DALYs with projections to 2030. *Int J Surg*. 2024;110:1951-67. DOI PubMed PMC
4. Smith JM, Kapur SK, Mericli AF, Baumann DP, Butler CE. Parastomal hernia repair. *Plast Aesthet Res*. 2022;9:10. DOI
5. Köckerling F, Sheen AJ, Berrevoet F, et al. The reality of general surgery training and increased complexity of abdominal wall hernia surgery. *Hernia*. 2019;23:1081-91. DOI PubMed PMC
6. Tuma F, Lopez RA, Varacallo MA. Anatomy, abdomen and pelvis: inguinal region (inguinal canal). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK470204/>. [Last accessed on 17 Apr 2025].
7. Roos MM, Bakker WJ, Schouten N, et al. Higher recurrence rate after endoscopic totally extraperitoneal (TEP) inguinal hernia repair with ultrapro lightweight mesh: 5-year results of a randomized controlled trial (TULP-trial). *Ann Surg*. 2018;268:241-6. DOI
8. Wu F, Zhang X, Liu Y, Cao D, Yu Y, Ma Y. Lightweight mesh versus heavyweight mesh for laparo-endoscopic inguinal hernia repair: a systematic review and meta-analysis. *Hernia*. 2020;24:31-9. DOI
9. Bakker WJ, Aufenacker TJ, Boschman JS, Burgmans JPJ. Heavyweight mesh is superior to lightweight mesh in laparo-endoscopic inguinal hernia repair: a meta-analysis and trial sequential analysis of randomized controlled trials. *Ann Surg*. 2021;273:890-9. DOI PubMed
10. Levy AS, Bernstein JL, Celie KB, Spector JA. Quantifying fascial tension in ventral hernia repair and component separation. *Hernia*. 2021;25:107-14. DOI PubMed
11. Dietz UA, Kudsi OY, Gokcal F, et al. Excess body weight and abdominal hernia. *Visc Med*. 2021;37:246-53. DOI PubMed PMC
12. de Beaux A, Becker S, Parent T, et al. Hernia Basecamp - a free to use, online hernia learning platform. Analysis of its use since launch in June 2021. *J Abdom Wall Surg*. 2023;2:11803. DOI PubMed PMC
13. Xu Y, Saiding Q, Zhou X, Wang J, Cui W, Chen X. Electrospun fiber-based immune engineering in regenerative medicine. *Smart Med*. 2024;3:e20230034. DOI PubMed PMC
14. Murphy BL, Ubl DS, Zhang J, Habermann EB, Farley DR, Paley K. Trends of inguinal hernia repairs performed for recurrence in the United States. *Surgery*. 2018;163:343-50. DOI PubMed
15. Palser TR, Swift S, Williams RN, Bowrey DJ, Beckingham IJ. Variation in outcomes and use of laparoscopy in elective inguinal hernia repair. *BJIS Open*. 2019;3:466-75. DOI PubMed PMC
16. Takeuchi M, Collins T, Lipps C, et al. Towards automatic verification of the critical view of the myopectineal orifice with artificial intelligence. *Surg Endosc*. 2023;37:4525-34. DOI
17. Takeuchi M, Collins T, Ndagijimana A, et al. Automatic surgical phase recognition in laparoscopic inguinal hernia repair with artificial intelligence. *Hernia*. 2022;26:1669-78. DOI
18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-44. DOI PubMed
19. Chan H, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. In: Lee G, Fujita H, editors. Deep learning in medical image analysis. Cham: Springer International Publishing; 2020. pp. 3-21. DOI PubMed PMC
20. De Coster M, Shterionov D, Van Herreweghe M, et al. Machine translation from signed to spoken languages: state of the art and challenges. *Univ Access Inf Soc*. 2024;23:1305-31. DOI
21. Cao H, Tan C, Gao Z, et al. A survey on generative diffusion models. *IEEE Trans Knowl Data Eng*. 2024;36:2814-30. DOI
22. Bengesi S, El-Sayed H, Sarker MK, Houkpati Y, Irungu J, Oladunni T. Advancements in generative AI: a comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEE Access*. 2024;12:69812-37. DOI
23. Liu C, Yu H. AI-empowered persuasive video generation: a survey. *ACM Comput Surv*. 2023;55:1-31. DOI
24. Ning Y, He S, Wu Z, Xing C, Zhang LJ. A review of deep learning based speech synthesis. *Appl Sci*. 2019;9:4050. DOI
25. ChatGPT. Available: <https://chatgpt.com>. [Last accessed on 17 Apr 2025].
26. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203-11. DOI PubMed
27. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv* 2016,

- arXiv:1603.04467. Available from: <https://doi.org/10.48550/arXiv.1603.04467>. [Last accessed on 17 Apr 2025]
28. Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch. 2017. Available from: <https://www.semanticscholar.org/paper/Automatic-differentiation-in-PyTorch-Paszke-Gross/b36a5bb1707bb9c70025294b3a310138aae8327a>. [Last accessed on 17 Apr 2025].
 29. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. *Nat Biomed Eng*. 2022;6:1346-52. DOI
 30. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *Lancet Digit Health*. 2021;3:e195-203. DOI PubMed
 31. Bohr A, Memarzadeh K. Artificial intelligence in healthcare. Academic Press, 2020. Available from: <https://shop.elsevier.com/books/artificial-intelligence-in-healthcare/bohr/978-0-12-818438-7>. [Last accessed on 17 Apr 2025].
 32. Panesar A. Machine learning and AI for healthcare: big data for improved health outcomes. Berkeley, CA: Apress, 2021. DOI
 33. Claus CMP, Cavaleiri M, Malcher F, et al. DECOMP Report: answers surgeons expect from an abdominal wall imaging exam. *Rev Col Bras Cir*. 2022;49:e20223172. DOI PubMed PMC
 34. Soler L, Mutter D, Marescaux J. Patient-specific anatomy: the new area of anatomy based on 3D modelling. In: Uhl J, Jorge J, Lopes DS, Campos PF, editors. Digital anatomy. Cham: Springer International Publishing; 2021. pp. 285-97. DOI
 35. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustainability*. 2021;13:1224. DOI
 36. Wasserthal J, Breit HC, Meyer MT, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell*. 2023;5:e230024. DOI PubMed PMC
 37. Zhang Q, Fu X, He K, et al. A new technique for the 3D reconstruction of the incisional hernia: a pilot study. *Clin Imaging*. 2020;67:91-4. DOI
 38. Yao S, Li JY, Liu FD, Pei LJ. Significance of measurements of herniary area and volume and abdominal cavity volume in the treatment of incisional hernia: application of CT 3D reconstruction in 17 cases. *Comput Aided Surg*. 2012;17:40-5. DOI
 39. Elhage SA, Deerenberg EB, Ayuso SA, et al. Development and validation of image-based deep learning models to predict surgical complexity and complications in abdominal wall reconstruction. *JAMA Surg*. 2021;156:933-40. DOI PubMed PMC
 40. Ayuso SA, Elhage SA, Zhang Y, et al. Predicting rare outcomes in abdominal wall reconstruction using image-based deep learning models. *Surgery*. 2023;173:748-55. DOI
 41. Kao AM, Arnold MR, Augenstein VA, Heniford BT. Prevention and treatment strategies for mesh infection in abdominal wall reconstruction. *Plast Reconstr Surg*. 2018;142:149S-55S. DOI PubMed
 42. Claus C, Furtado M, Malcher F, Cavazzola LT, Felix E. Ten golden rules for a safe MIS inguinal hernia repair using a new anatomical concept as a guide. *Surg Endosc*. 2020;34:1458-64. DOI PubMed
 43. Lin C, Zhu Z, Zhao Y, Zhang Y, He K, Zhao Y. SGT++: improved scene graph-guided transformer for surgical report generation. *IEEE Trans Med Imaging*. 2024;43:1337-46. DOI
 44. Nagaya T, Nakamura YA, Choyke PL, Kobayashi H. Fluorescence-guided surgery. *Front Oncol*. 2017;7:314. DOI
 45. Diana M. Enabling precision digestive surgery with fluorescence imaging. *Transl Gastroenterol Hepatol*. 2017;2:97. DOI PubMed PMC
 46. Barberio M, Benedicenti S, Pizzicannella M, et al. Intraoperative guidance using hyperspectral imaging: a review for surgeons. *Diagnostics*. 2021;11:2066. DOI PubMed PMC
 47. Yoon J. Hyperspectral imaging for clinical applications. *BioChip J*. 2022;16:1-12. DOI
 48. Aasvang EK, Meyhoff CS. The future of postoperative vital sign monitoring in general wards: improving patient safety through continuous artificial intelligence-enabled alert formation and reduction. *Curr Opin Anaesthesiol*. 2023;36:683-90. DOI PubMed
 49. Tanner J, Rochon M, Harris R, et al. Digital wound monitoring with artificial intelligence to prioritise surgical wounds in cardiac surgery patients for priority or standard review: protocol for a randomised feasibility trial (WISDOM). *BMJ Open*. 2024;14:e086486. DOI PubMed PMC
 50. Le DTP, Pham TD. Unveiling the role of artificial intelligence for wound assessment and wound healing prediction. *Explor Med*. 2023;4:589-611. DOI
 51. Ganesan O, Morris MX, Guo L, Orgill D. A review of artificial intelligence in wound care. *Art Int Surg*. 2024;4:364-75. DOI
 52. Jamadar DA, Jacobson JA, Girish G, et al. Abdominal wall hernia mesh repair: sonography of mesh and common complications. *J Ultrasound Med*. 2008;27:907-17. DOI
 53. Lacour M, Ridereau Zins C, Casa C, et al. CT findings of complications after abdominal wall repair with prosthetic mesh. *Diagn Interv Imaging*. 2017;98:517-28. DOI
 54. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering*. 2023;10:1435. DOI PubMed PMC
 55. Sajdeya R, Narouze S. Harnessing artificial intelligence for predicting and managing postoperative pain: a narrative literature review. *Curr Opin Anaesthesiol*. 2024;37:604-15. DOI PubMed
 56. van de Sande D, van Genderen ME, Verhoef C, et al. Optimizing discharge after major surgery using an artificial intelligence-based decision support tool (DESIRE): an external validation study. *Surgery*. 2022;172:663-9. DOI
 57. Geoghegan L, Scarborough A, Wormald JCR, et al. Automated conversational agents for post-intervention follow-up: a systematic review. *BJS Open*. 2021;5:zrab070. DOI PubMed PMC
 58. Park JJ, Tiefenbach J, Demetriades AK. The role of artificial intelligence in surgical simulation. *Front Med Technol*. 2022;4:1076755.

[DOI](#) [PubMed](#) [PMC](#)

59. Bencteux V, Saibro G, Shlomovitz E, et al. Automatic task recognition in a flexible endoscopy benchtop trainer with semi-supervised learning. *Int J Comput Assist Radiol Surg.* 2020;15:1585-95. [DOI](#)
60. Eckhoff JA, Rosman G, Altieri MS, et al. SAGES consensus recommendations on surgical video data use, structure, and exploration (for research in artificial intelligence, clinical quality improvement, and surgical education). *Surg Endosc.* 2023;37:8690-707. [DOI](#) [PubMed](#) [PMC](#)
61. Mutter D, Vix M, Dallemagne B, Perretta S, Leroy J, Marescaux J. WeBSurg: an innovative educational Web site in minimally invasive surgery - principles and results. *Surg Innov.* 2011;18:8-14. [DOI](#) [PubMed](#)
62. Seenivasan L, Islam M, Krishna AK, Ren H. Surgical-VQA: visual question answering in surgical scenes using transformer. *arXiv* 2022, arXiv:2206.11053. Available from: <https://doi.org/10.48550/arXiv.2206.11053>. [Last accessed on 17 Apr 2025]
63. Browne R, Gull K, Hurley CM, Sugrue RM, O'Sullivan JB. ChatGPT-4 can help hand surgeons communicate better with patients. *J Hand Surg Glob Online.* 2024;6:436-8. [DOI](#) [PubMed](#) [PMC](#)
64. Ruksakulpiwat S, Kumar A, Ajibade A. Using ChatGPT in medical research: current status and future directions. *J Multidiscip Healthc.* 2023;16:1513-20. [DOI](#) [PubMed](#) [PMC](#)
65. Lai H, Ge L, Sun M, et al. Assessing the risk of bias in randomized clinical trials with large language models. *JAMA Netw Open.* 2024;7:e2412687. [DOI](#) [PubMed](#) [PMC](#)
66. Kirova VD, Ku CS, Laracy JR, Marlowe TJ. Software engineering education must adapt and evolve for an LLM environment. In: *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, New York, USA. 2024. Association for Computing Machinery; 2024. pp. 666-72. [DOI](#)
67. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245-7. [DOI](#) [PubMed](#) [PMC](#)
68. Daes J, Felix E. Critical view of the myopectineal orifice. *Ann Surg.* 2017;266:e1-2. [DOI](#) [PubMed](#)
69. Barberio M, Collins T, Bencteux V, et al. Deep learning analysis of in vivo hyperspectral images for automated intraoperative nerve detection. *Diagnostics.* 2021;11:1508. [DOI](#) [PubMed](#) [PMC](#)
70. Wang C. Calibration in deep learning: a survey of the state-of-the-art. *arXiv* 2023, arXiv:2308.01222. Available from: <https://doi.org/10.48550/arXiv.2308.01222>. [Last accessed on 17 Apr 2025]
71. Borys K, Schmitt YA, Nauta M, et al. Explainable AI in medical imaging: an overview for clinical practitioners - Saliency-based XAI approaches. *Eur J Radiol.* 2023;162:110787. [DOI](#)
72. Alberdi E, Povykalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Acad Radiol.* 2004;11:909-18. [DOI](#) [PubMed](#)