

MEDICAL ROBOTS

Will your next surgeon be a robot? Autonomy and AI in robotic surgery

Samuel Schmidgall¹, Justin D. Opfermann², Ji Woong Kim^{2†}, Axel Krieger^{2*}

State-of-the-art surgery is performed robotically under direct surgeon control. However, surgical outcome is limited by the availability, skill, and day-to-day performance of the operating surgeon. What will it take to improve surgical outcomes independent of human limitations? In this Review, we explore the technological evolution of robotic surgery and current trends in robotics and artificial intelligence that could lead to a future generation of autonomous surgical robots that will outperform today's teleoperated robots.

INTRODUCTION

Every year, more than 150,000 patients in the US are diagnosed with colon cancer (1), and laparoscopic colectomy stands as their sole curative treatment. However, the surgery's success varies substantially because of differences in surgeons' skills, experience, and techniques, leading to complication rates reported as high as 23% in some studies (2). Despite the advancements in teleoperated robotic surgery, it has not notably reduced the risk of complications such as anastomotic leaks (3). Given these challenges, there is a growing interest in surgical robots that can perform procedures autonomously. These robots have the potential for both greater consistency and precision than even the most skilled surgeons, potentially lowering complication rates and decreasing patient wait times. Moreover, they could extend essential surgical care to areas lacking expert surgeons, including remote locations on Earth and space missions, ultimately democratizing access to quality surgical interventions.

This Review explores the technological evolution of robotic surgery and current trends in robotics and artificial intelligence (AI) that could lead to a future generation of autonomous surgical robots that will outperform today's teleoperated robots. We expand on previous reviews in the field that focused on clinical applications (4), regulatory and ethical considerations (5), medical robotics retrospectives (6), spine surgery (7), glossary terms (8), or health ecosystems (9) and those that occurred well before the current wave of AI in robotics (10–13). In this work, we briefly explore the evolution of surgical techniques from ancient times to the present day, highlighting key milestones that have shaped modern surgery. We then discuss the current state of robotic-assisted surgery (RAS) and its limitations, setting the stage for the potential of autonomous surgical robots. We examine the levels of surgical robot autonomy, from teleoperation to full autonomy, and discuss the AI methods driving progress toward increased surgical autonomy. It is important to note that this Review is limited to end-to-end learning methods, such as reinforcement learning and imitation learning, which have shown promising results for achieving autonomous surgical actions. We do not cover broader AI categories—including model-based methods like task planning and motion planning—in depth (13–15), because our goal is to highlight the recent advances in learning-based surgical autonomy.

Specifically, we focus on two primary approaches: simulation and imitation-based learning. We analyze the advantages and challenges of each method, including the development of sophisticated surgical simulators and the application of reinforcement learning and imitation learning techniques. Throughout the Review, we consider the potential impact of autonomous surgical robots on health care delivery, patient outcomes, and the future of surgical practice. We conclude with a perspective on the current trends in autonomous robotics that will likely shape the future of robotic surgery.

An overview of surgical techniques

Surgical approaches

Despite the history of surgery dating as far back as 3000 BCE (16), it was not until the 19th century CE that the innovation of anesthesia and aseptic techniques would lead to a paradigm shift in surgical care. Sedation, sterilization, and sterile techniques reduced infection rates and enabled more complex surgeries. Over the course of the past 60 years, surgeries have evolved from a largely open technique to procedures with no incision at all (Fig. 1). Open surgery refers to those procedures where large incisions are used to expose entire organs and allow the surgeon to directly manipulate the surgical field, such as in open-heart surgery and microvessel reconstruction. Open surgeries were made less invasive by the invention of laparoscopes and long surgical tools that provide visualization and manipulation of tissue through small surgical ports, known as laparoscopic surgery (17). These procedures are referred to as single-port surgeries when only a single access point is used. Surgical incisions can be completely eliminated when the single-port surgery is performed through a natural orifice, a procedure called natural orifice transluminal endoscopic surgery (NOTES) (18, 19). NOTES potentially reduces postoperative pain, shortens recovery times, and minimizes the risk of wound-related complications. However, this technique presents unique challenges, including the need for specialized instruments, advanced endoscopic skills, and careful consideration of infection control measures. The development of specialized robotic systems tailored for single-port surgeries is expected to reduce these barriers and expand their application across a broad range of surgical interventions (20, 21).

Robotic-assisted surgery

As surgical approaches become progressively less invasive from open to NOTES procedures, the technical skill and surgical training required to achieve high-quality surgical outcomes drastically increase. The tools used in laparoscopic surgery, for instance, may only have a single degree of freedom, which makes tissue cutting

¹Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA. ²Department of Mechanical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA.

*Corresponding author. Email: axel@jhu.edu

†Present address: Stanford University, Stanford, CA, USA.

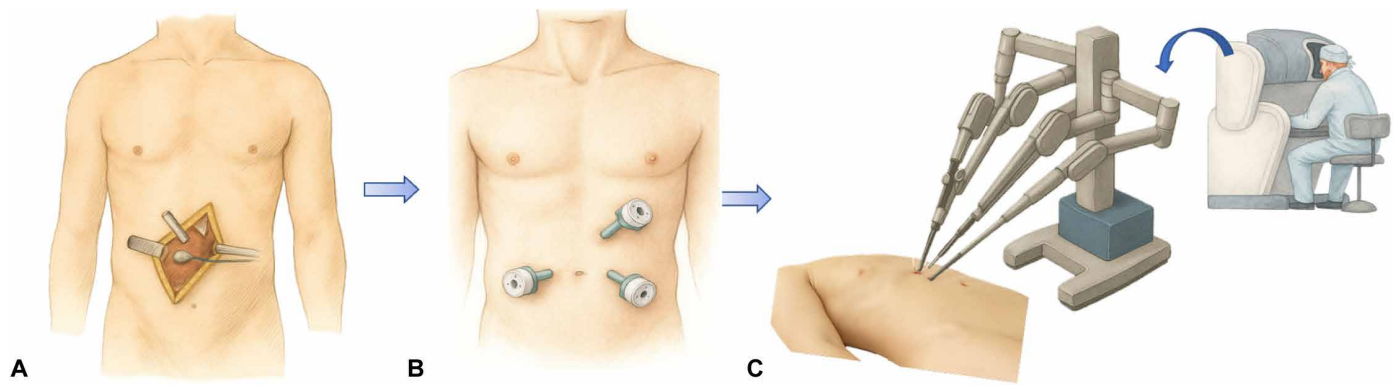


Fig. 1. The evolution of surgical techniques. (A) Surgeries began with open interventions where internal organs are exposed with a large incision. (B) Clinical and technical innovations led to less-invasive techniques where tools were inserted through multiple ports to conduct laparoscopic surgeries. (C) Today's state-of-the-art surgery is performed by human-controlled robotic surgical systems using multi- and single-port approaches.

and suturing challenging for even the most experienced surgeon. These challenges are amplified in single-port and natural orifice surgeries where the working space is further constrained, resulting in tool collisions and small working volumes that restrict surgical motions. Minimally invasive surgery is also notorious for poor ergonomics because of the surgeon's posture, whereas visualization is limited because of two-dimensional (2D) views of the surgical field.

Today's surgical robots are specifically designed to restore the surgeon's surgical dexterity when manipulating tissue and tools while also improving the ergonomics and visualization of minimally invasive procedures (Fig. 1C). These systems use a "leader" and "follower" design whereby a surgeon sits at an operative console (leader) and controls the exact motions of a robotic arm and tool located at the patient's bedside (follower). The tools used in robotic surgery have multiple degrees of freedom at the distal end, which restore dexterity to the surgeon and enable wrist motions of the tools, creating the perception that the surgeon's hands are directly manipulating the distal end of the surgical instrument within the patient. The surgeon's positioning at the console also improves surgical ergonomics because the head and neck rest in a natural position to observe intraoperative video from the surgical camera. The displayed endoscopic video is stereoscopic, providing qualitative 3D visualization and restoring depth perception to the surgeon. Today, RAS is commonly performed for laparoscopic, single-site, and natural orifice procedures. With more than 90% of radical prostatectomies performed with a robotic approach, RAS is the standard of care for urology, but global surgical trends illustrate rapid adoption of RAS in general surgery and gynecology (22). In 2023, more than 2.2 million surgeries were performed with da Vinci surgical platforms alone, and global cases are expected to top 2.5 million by the end of 2024 (23).

THE NEED FOR AUTONOMOUS ROBOTIC SURGERY

Although RAS has improved ergonomics and introduced technical advantages, its cost-effectiveness and long-term outcome benefits remain subjects of debate (24, 25). Current surgical paradigms rely on the surgeon to control every motion of the robot, and so surgical outcomes are still directly related to the surgeon's training, skill, and decision-making during a procedure. There is a substantial learning curve in robotic surgery (26), surgical outcomes vary greatly between hospitals (27), and postoperative complications differ by surgeon (28).

Surgical robots, which have built-in autonomy, present a promising solution to these challenges, offering the potential to improve efficiency, safety, and consistency in surgical procedures. By providing expert-level care independent of location or surgeon experience, these systems could substantially reduce complication rates in complex operations, such as kidney and bowel surgeries. Moreover, they have the potential to address health care disparities by extending high-quality surgical care to underserved areas, contain rising health care costs, and protect health care workers during pandemics. This becomes increasingly important as surgeon shortages continue to increase across the world (29).

The successful integration of autonomous robots could revolutionize surgical practice in several ways. These systems could support surgeons by handling repetitive tasks, enable simultaneous oversight of multiple operations, address projected surgeon shortages (particularly in developing countries), and improve overall surgical safety and consistency. However, the path to fully autonomous surgery is not without obstacles. Key challenges include accurately differentiating between healthy and diseased tissues in complex, variable surgical environments; maintaining robust perception when blood, surgical smoke, fog, or tissue debris occlude the camera (30–32); compensating for rapid lighting changes caused by endoscope repositioning or electrocautery flashes; accommodating patient-specific anatomical variations; managing real-time tissue deformations during procedures [e.g., bowel wall excursions of ± 8 mm at 1 Hz due to respiration and peristalsis (33)]; and reconstructing stable, drift-free 3D models of constantly deforming anatomy.

Navigating the complexities of soft tissue surgery presents additional hurdles where preoperative planning is less feasible because of variability and access limitations. To overcome these challenges, autonomous surgical robots will require advanced capabilities in accurate imaging, adaptive planning, and precise execution. In the following sections, we will introduce the varying levels of autonomy (LoAs), as well as discuss the technical paths using AI that have the potential for furthering autonomy in robotic surgery.

LEVELS OF SURGICAL ROBOT AUTONOMY

The integration of robotic surgical systems, augmented with AI, 3D visualization, and specialized tools, has introduced autonomous

capabilities surpassing those of expert surgeons in both synthetic and living tissues, with some aspects already adopted in clinically approved systems. To categorize these advancements, six tiers of LoAs were introduced in 2017 (5) and then refined in 2019 (10) to better define human supervision requirements and task-specific performance metrics. We can further delineate and clarify differences in the LoAs by adding a secondary metric originally proposed by IEC/TR 60601-4-1 (International Electrotechnical Commission Technical Report; International Organization for Standardization standard for the guidance and interpretation of medical electrical equipment and medical electrical systems) and interpreted by Kaber and Endsley (34) called the degree of autonomy (DoA) (Table 1). The DoA classifies the relationship between human and robot interaction during four different operating functions that can be defined in the context of surgery as follows: monitor—collect information about the patient, operator, or surgical task; generate—create possible plans or routines to achieve a surgical goal based on monitoring; select—choose a plan or routine to execute; and execute—carry out the selected plan or routine. In this Review, we differentiate the LoAs of surgical robotic systems on the basis of these four operating functions, because they provide more granularity to better capture the combinations of human and robot decision-making that constitute the autonomous capability of each specific system.

No autonomy (L-0)

With no autonomy, the human surgeon is solely responsible for monitoring, generating, selecting, and executing the surgery. During the surgery, no active robotic equipment is used. No autonomy is the current standard of care for many surgical centers that do not have access to robotic surgical equipment and is commonly referred to as manual surgery. Robotic systems for capsule endoscopy that only enable the operator to visually inspect the surgical scene and not provide intervention, such as the NaviCam (35), fall into this category as well.

Robot assistance (L-1)

At this level, monitoring may be shared between the human and robot, whereas the human operator retains control over the generation, selection, and execution of surgical plans. The robot supports low-level mechanical operations to improve human performance without direct involvement in surgical tasks. Examples include teleoperation with systems like the da Vinci surgical robot (36–38), Senhance Surgical System (39), or MUSA-3 robot (40), which offer dexterous tool control and tremor filtering. A key to their success is cable-driven actuators enabling wristed tool motions, adding two degrees of freedom for a full six degrees of motion inside the body, intuitively mapped to replicate the surgeon’s hand movements. Other robotic

Downloaded from https://www.science.org at The Hong Kong University of Science and Technology (Guangzhou) on May 25, 2026

Table 1. Classifying autonomous surgical techniques on the basis of their respective LoAs and DoAs. H represents human actions taken; R represents robot actions taken.

LoA	Description	DoA	Description	Monitor	Generate	Select	Execute
0	No autonomy	0	Full manual: There is no robotic interaction during the surgical procedure, and the human is responsible for all surgical planning, decision making, and execution.	H	H	H	H
1	Teleoperation	1	Teleoperation: The robot assists the human operator during surgery, but the human maintains continuous control of the robot. The robot may perform mechanical tasks to enhance the operator’s performance, but it does not perform surgical tasks on the patient.	H/R	H	H	H
2	Task autonomy	2	Preprogrammed execution: The human generates and selects the surgical plan to be performed, and the robot executes that plan.	H/R	H	H	R
		3	Shared decision: Both the human and robot generate possible surgical plans, but the human has the sole decision to select the surgical plan to be executed. The human and robot share execution of the plan.	H/R	H/R	H	H/R
		4	Decision support: The robot generates surgical plans that the human may choose to augment. The robot then performs execution of the surgical plan.	H/R	H/R	H	R
3	Supervised autonomy	5	Blended decision: The human and robot share in the generation and selection of the surgical plan. However, the robot makes an initial selection that the human can choose to confirm or modify.	H/R	H/R	H/R	R
		6	Guided decision: The robot is responsible for generating all of the surgical plans, and the human selects the surgical plan. The robot executes the surgical plan without human assistance.	H/R	R	H	R
4	High-level autonomy	7	Autonomous decision: The robot generates, selects, and executes a surgical plan. The surgical plan can be augmented by the human before it is selected or executed.	H/R	H/R	R	R
		8	Operator monitoring: The robot generates, selects, and executes the surgical plan without human assistance. The human can only monitor the task execution and intervene in cases of emergency.	H/R	R	R	R
5	Full autonomy	9	Full autonomy: The robot executes all steps of a surgical procedure and does not require monitoring from a human.	R	R	R	R

aids include flexible endoscope holders (41), the EndoAssist laparoscope holder (42), the Jaimy needle grasper (43), the Edison system for histotripsy (44), the MagnetoSuture anastomosis robot (45), and Micron (46) for stabilizing movements in delicate surgeries.

Task autonomy (L-2)

With task-level autonomy, the human performs the selection of a surgical plan, whereas monitoring, generation, and execution of low-risk tasks may be shared with the robot, shifting from continuous to discrete surgeon control. The human's selection of the final plan and shared control during its execution are defining factors of L-2 systems. These plans can range from low-level motions, like needle grasping and insertion, to complex tasks, such as anastomosis. Examples include the AutoLap for camera positioning (47), the MAKO robot for arthroplasty (48), the KINEVO 900 for microscope positioning (49), the Artis Zeego for targeted x-ray imaging (50), the Ion endoluminal system for bronchoscopy (51), and research systems for needle steering (52) (considered L-3 when autonomous). Research with the da Vinci Research Kit (dVRK) has demonstrated L-2 autonomy for laboratory-based debridement (53) and blood clearing (54). The success of L-2 systems is largely enabled by advancements in image fusion and registration between preoperative imaging and the robot. Furthermore, robust software for tool tracking via color or fiducial segmentation and algorithms for differentiating background from foreground tissues have advanced intraoperative applications.

Supervised autonomy (L-3)

With L-3 autonomy, the human and robot share monitoring and generation of surgical plans, with the robot potentially selecting the final plan. The defining characteristic of L-3 systems is their ability to execute a task without human oversight or intervention, a substantial shift in responsibility from L-2 systems. For instance, the Smart Tissue Autonomous Robot (STAR) (55) generates multiple anastomosis plans and autonomously executes the one selected by a surgeon. The critical technical enabler for the STAR was quantitative 3D endoscopic imaging, allowing intraoperative 3D reconstruction of the surgical field to overlay plans with submillimeter accuracy. In addition, machine learning for real-time tissue tracking allows the STAR to compensate for tissue motion. Similarly, autonomous needle steering with (52) operates as a blended decision L-3 system. Clinical systems like the TSolution One (56) also demonstrate this level by performing surgical drilling routines on the basis of preoperative plans.

High-level autonomy (L-4)

High-level autonomy systems have shared human and robot monitoring and surgical plan generation, but selecting and executing surgical plans are the responsibility of the robot. The defining characteristic of L-4 systems is that once a procedure begins, decision-making is performed by the robot, including potential deviations from preselected surgical plans. L-4 systems have minimal human intervention, which is commonly limited to monitoring the surgery and intervening in emergency situations. The CyberKnife (57) and the experimental Veebot (58) system represent this, performing tasks like radiosurgery for deep tissue tumor ablation and vein cannulations that autonomously follow a preoperative plan, whereas the VisuMax robotic system is used for corneal refractive surgery (59). Most clinical L-4 systems are possible because the procedure does not require tissue

dissection to expose the diseased state, and treatment can be applied noninvasively to a targeted region. This approach eliminates technical considerations and complications in automation that arise from soft tissue deformations during surgery, which has simplified automating the procedure.

Full autonomy (L-5)

With full autonomy, robots could complete surgeries independently, even in unforeseen circumstances, a stage not yet realized. The robot would solely perform monitoring, generation, selection, and execution of surgical plans without human intervention. This requires advanced detection, processing, and response capabilities, particularly for soft tissue surgeries. Although a path has been envisioned (60), much work remains before L-5 autonomy can be developed.

In summary, a delineation exists for procedural suitability for higher autonomy. Tasks involving rigid structures (orthopedics), noninvasive energy delivery from preoperative plans (radiosurgery), or structured subtasks (supervised anastomosis) are most amenable to near-term levels 3 or 4 autonomy. Conversely, procedures with complex, unpredictable soft tissue interactions, especially with deformation or perception challenges from bleeding or smoke, require continued human oversight (level 1 or 2) or supervisory roles (level 3 or 4). Bridging this gap for complex soft tissue surgery is a primary long-term goal, contingent on major advancements in sensing, planning, and adaptive control.

Autonomy and federal approval

Although the integration of autonomy in surgical robots has made progress in recent years, the regulatory frameworks by the US Food and Drug Administration (FDA) have not progressed at the same rate to match these advancements. A recent systematic review (61) indicated that most surgical robots approved by the FDA remain at a basic LoA, requiring direct control from surgeons. The analysis shows that although some surgical robots have achieved conditional autonomy (level 3), most are still at level 1, demonstrating a gap between autonomous capability and regulatory classifications. This delay suggests a need for updated regulatory frameworks that can more accurately manage the increasing autonomy in surgical robotics, especially with the introduction of autonomous AI-based robotic systems.

AI-based surgical autonomy

Figure 2 shows that less invasive robotic systems are inherently less autonomous because of the technical challenges of minimizing surgical invasiveness. Current model-based approaches are reaching their limits, illustrating the need for a paradigm shift to increase surgical autonomy for these procedures. Bridging this gap and shifting the trend line upward require new surgical paradigms.

Intraoperative planning and control must meet millisecond-level deadlines while accommodating large, nonlinear tissue motions. For example, laparoscopic cameras can pivot $>90^\circ$, and instrument tips can exceed 60 mm s^{-1} . Soft-tissue stiffness may vary by two orders of magnitude, and respiration can shift targets by up to 30% of their diameter. Classical path planners like rapidly exploring random tree (RRT) assume static, rigid workspaces, incurring 3- to 7-mm positional errors when tissue elasticity or kinematic constraints are introduced. Although promising, recent model-predictive controllers require $>50 \text{ ms}$ per update, an order of magnitude slower than the $<5\text{-ms}$ control loop needed for safe suturing. Closing this

The Trend of Autonomy and the Surgical Approach

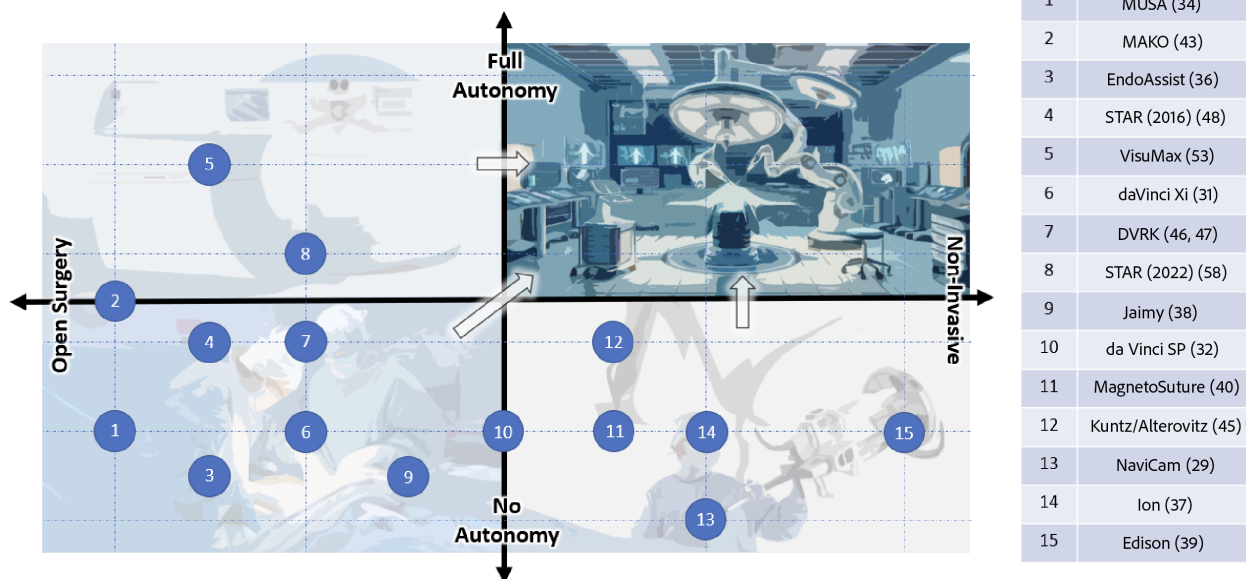


Fig. 2. Illustrating the trend of autonomy of well-known surgical robotics as a function of the surgical approach. The x axis represents the measure of invasiveness of the surgery, and the y axis represents the amount of robotic autonomy. In general, as surgical approaches become less invasive, the need for robotic solutions increases, but the autonomous capability of those systems has not been realized. This is illustrated by the absence of highly autonomous and noninvasive robots in the upper right quadrant. Developments in AI are needed to fundamentally shift the development of robots such that less-invasive surgical approaches can be performed with increased LoAs. Each number represents a notable robotic surgical system plotted at their approximate LoA and invasiveness.

gap demands hybrid methods that fuse fast, sampling-based planners with neural surrogates of soft-body dynamics and closed-loop feedback.

Historically, the greatest advancements in surgical autonomy have been model-based (55, 62–64); however, AI now offers a promising direction to learn directly from expert surgeons and from simulated surgical scenes. These methods may be the key to increasing autonomy in less invasive procedures. In the next sections, we outline the key methods toward increasing surgical autonomy with AI and where the future of AI in robotic surgery is moving. Although AI plays a pivotal role in the future of surgical robotics, it is important to recognize that it is not the sole factor driving innovation in surgical robotics. Many other works, such as smaller, softer robotic systems and improved image guidance, contribute to the state of the art and encompass a broad range of interdisciplinary efforts.

AI METHODS TOWARD INCREASED SURGICAL AUTONOMY

Advances in AI are expanding the autonomy of surgical systems, although they are not yet ready to replace expert surgeons. This Review focuses on emerging learning-based AI approaches—namely, reinforcement learning (RL) and imitation learning (IL)—because these methods show great promise in achieving higher LoAs (see Fig. 3). We also explore future trends, highlighting the potential for high-capacity vision-language-action (VLA) models (65). Although our Review emphasizes learning-based methods, we recognize the contributions of non-learning-based approaches to robotic autonomy, including task planning, motion planning, and computer vision.

Overview of reinforcement learning

RL (66) is a machine learning framework in which a robotic agent learns how to act in an environment by receiving rewards for its actions. Key concepts in RL include the state, action, reward, and policy. A state (denoted s) represents the environment at a given moment—for instance, an endoscopic image or the current joint positions of a surgical robot. An action (denoted a) refers to a possible operation the agent can take while in a given state, such as moving a surgical instrument to grasp tissue. A reward (denoted r) is a numerical signal that reflects the quality of the agent's action in a specific state—for example, how close the robot is to a surgical target. Last, a policy (denoted π) is a strategy that maps states to actions, guiding the agent in selecting which action to take next on the basis of the current state.

For example, imagine a robot learning to position a needle in a tissue sample. The state could be the robot's camera view, the action could be different ways of moving the instrument, and the reward might increase when the robot properly aligns the needle for a precise insertion.

In RL, the behavior of the agent is dictated by a policy $\pi: \pi(a|s) = \Pr(a|s)$, which means that for any given state s , the policy assigns a probability to each possible action a . In contrast, the environment's state transition probabilities, denoted by $P(s'|s, a)$, define the likelihood of transitioning to a new state s' given the current state s and the action a taken.

Value function and Q-function

The value function, denoted $V^\pi(s)$, estimates how much total reward the robot can gain starting from state s under policy π . The Q-function, $Q^\pi(s, a)$, refines this idea further by estimating the total reward for

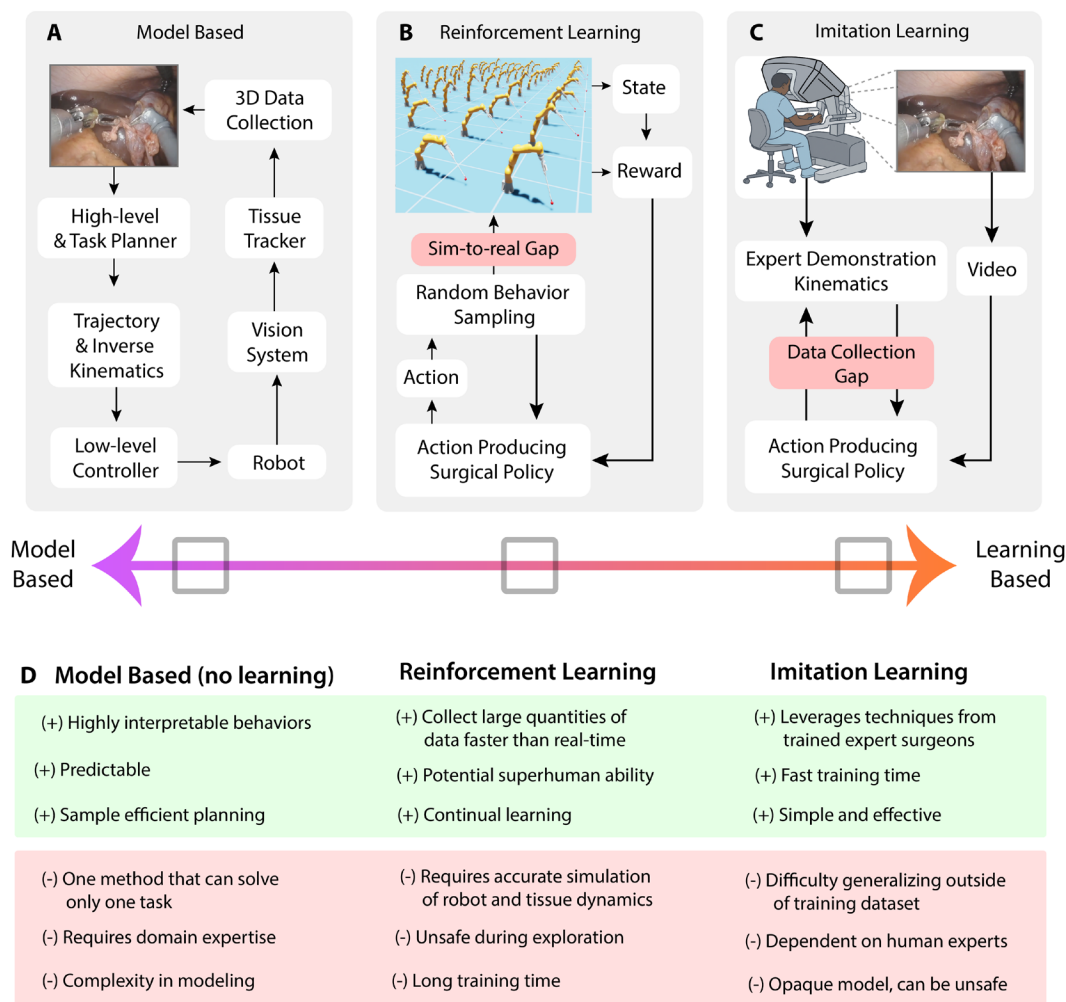


Fig. 3. Comparison of three major control techniques in robotic surgery. (A) Demonstration of model-based workflow. (B) RL workflow, with the major challenge being translating learned policy from simulation to real hardware. (C) IL workflow, with the main challenge being collecting data to perform imitation. (D) Comparison of the positive and negative factors of each technique.

taking a specific action a in state s (67). These functions guide the robot in choosing actions that ultimately yield higher accumulated rewards.

Q-learning

A well-known RL method is Q-learning, which can learn an optimal policy without needing a detailed model of the environment (68). It updates its Q-function on the basis of the difference between the predicted reward and the actual reward received when an action is taken. Because Q-learning can learn from different data sources, it is often described as an “off-policy” algorithm. In robotic surgery, Q-learning could iteratively improve how an agent handles specific subtasks, such as maneuvering instruments around delicate tissues. For a surgical robot, this is particularly beneficial because accurately modeling the complex interactions in a surgical setting (such as tissue deformation, tool-tissue interactions, and bleeding) is extremely challenging. Instead, the robot learns effective policies directly from its interactions (or simulated interactions) with the environment, thereby adapting to the variability and uncertainty inherent in surgical procedures.

Surgical simulators

The development of surgical simulations has increased in recent years, with most of these simulators focusing on surgical skill tasks, such as picking up pegs or transferring a needle. However, more advanced simulators are just beginning to emerge, more closely resembling real surgical scenes (69–71). Simulators provide a safe environment for experimentation for robots. This is especially important for surgical robots, where the robot is operating inside real patients who could be harmed if any imprecise actions are taken.

One of the pioneering platforms in this area is the da Vinci Reinforcement Learning (dVRL) platform (72). This platform was designed to help engineers without a surgical background create algorithms for autonomous surgery. dVRL offers two basic training tasks: a robotic reach task and a pick-and-place task. A controller trained using dVRL was demonstrated to transition from simulation to actual robotic hardware, using a robot to suck blood in a surgical demonstration. Built on the Unity3D platform, which supports deformable objects, UnityFlexML (73) also demonstrates successful policy transfer to the dVRK, similar to dVRL.

LapGym (74) is an RL environment that offers a diverse range of tasks for developing and testing automation in laparoscopic surgery. This platform covers skills like spatial reasoning, manipulating and grasping deformable objects, dissection, and thread manipulation. It presents a variety of image-based learning challenges and supports multiple data formats like RGB (red, green, blue), depth, point clouds, and semantic segmentation. CRESSim builds on PhysX (75) and enables the simulation of several surgical features, such as soft tissue and body fluids (70), focusing on photorealism. SurRoL (76) is an open-source RL-focused simulation platform, compatible with the dVRK like its predecessors. It provides environments for different components of the da Vinci system and multiple arms and has been shown to transfer to real-world dVRK problems, such as simple pick-and-place and point tracking. Last, Surgical Gym (69) and ORBIT-Surgical (71) are open-source graphical processing unit (GPU)-based RL environments that support several surgically relevant training tasks. These works focus on making the simulator more efficient, which made training data much more accessible than previously possible using a GPU-based physics simulator. Overall, this platform (69) demonstrates between 100 and 5000 times faster training times compared with previous surgical learning platforms (70, 76). Simulation-based RL methods provide rapid policy development by enabling the exploration of a wide range of scenarios in a safe, virtual environment; however, the sim-to-real gap remains a challenge because of simplified physical dynamics and limited photorealism.

Challenges with simulation-based learning

The simulation-based RL approach has many challenges that prevent it from being used in certain fields, particularly in surgery. One of the greatest challenges thus far is the current design of simulators. Now, most simulators are also not capable of capturing properties that are reflective of the real world in a phenomenon referred to as the sim-to-real gap (77–79).

The sim-to-real gap

Although simulations can be highly detailed and realistic, they inevitably cannot match certain aspects of real-world complexity. This gap leads to models that perform well in simulation but fail to generalize to real-world conditions (78–80). There are two primary ways in which simulators will not match reality: physical dynamics and photorealism (80, 81). Physical dynamics in simulations are generally a simplified version of real-world physics. Elements like tissue elasticity, complex tool-tissue interactions, and bleeding are not accurately captured. The visual accuracy, or photorealism, of simulations also falls short of real-world environments (81). Fluid light reflection, tissue texture, and bleeding under varying conditions are difficult to replicate precisely (33). This is particularly problematic for tasks that require precise visual input, such as complex manipulation tasks encountered during surgery.

Some works have focused on crossing the sim-to-real gap with surgical robots for physical dynamics, such as for grasping tissue (82), online registration (83), and pose estimation (84). Other works have considered using intermediate visual representations, such as segmentation maps, which do not vary substantially between simulation and reality, to perform blood suctioning tasks (33). A few works have achieved direct simulation-to-real hardware transfer using RGB visual inputs for rigid (85) and soft-body (80) manipulation tasks.

Imitation-based autonomy

Imitation-based autonomy in surgical robotics takes inspiration from how human surgeons learn and refine their skills. Just as junior surgeons observe and emulate the techniques of experienced surgeons, models can also be trained to mimic the actions of expert surgeons. This approach involves capturing the movements, decision-making processes, and techniques of skilled surgeons as they perform the procedures.

One approach to this technique is IL (86–88). In IL, robots are trained on datasets of recorded surgical procedures by expert surgeons. These datasets contain rich information about tool movements, tissue interactions, and decision points throughout various surgical tasks (89–91). By learning from these data, robotic systems can develop policies that aim to replicate expert-level performance across a range of surgical scenarios (86).

One of the main advantages of imitation-based over reward-based learning is that this method can often bypass the need for a highly accurate simulator, thereby avoiding sim-to-real issues. In addition, IL can potentially adapt to different surgical styles, because it can be trained on demonstrations from multiple users with diverse behaviors (92). However, this approach also faces challenges, such as the need for large, high-quality datasets of surgical demonstrations and the potential to replicate suboptimal behaviors if present in the training data. In the next section, we provide background of two major techniques in IL and discuss specific examples from prior work.

Overview of imitation learning

IL can be particularly valuable in surgical contexts where defining explicit reward signals is challenging because the desired behavior is complex. IL can be considered a form of supervised learning, with the surgeon acting as the expert, which the model must learn to imitate.

The core concept of IL involves learning a policy that maps states to actions. The states can often be defined as visual observations of the surgical scene (e.g., images from an endoscope) or sensorimotor information (e.g., current effector pose of the robot arms) (86, 93). Actions can often be defined as desired goal poses that the end effectors must reach (86, 93). The primary objective is to develop a policy that accurately replicates the expert surgeon's decision-making process on the basis of a dataset of state-action pairs. There are two main approaches to IL: direct imitation, which learns to replicate the expert's policy by outputting similar commands given the same state, and indirect imitation, which learns a reward function that captures the expert's behavior.

In IL, the learning objective is typically formalized as minimizing a loss function. A common example is the mean-squared error, which measures the difference between the learned policy's output and the expert's action for each state in the demonstration dataset (86, 93).

Behavior cloning

Behavior cloning (BC) is a direct approach to IL that closely aligns with supervised learning principles. It involves training a policy function from a dataset of expert demonstrations. The optimization problem in BC aims to minimize the difference between the learned policy's actions and the expert's actions across all demonstrations.

BC faces several challenges in surgical robotics. The trained policy may encounter states not present in the training set, leading to potentially unpredictable behavior. This is particularly problematic

when the demonstrations come from sequential states and actions, resulting in a mismatch in distributions between the learned policy and expert demonstrations (94). Creating a comprehensive dataset often requires manually collecting data over many possible set of states, which can be labor intensive.

To address this challenge, a common approach is to implement some form of DAgger (Dataset Aggregation) (94–96), a procedure where the operator corrects the policy mistakes at the time of deployment and the corrections become part of the training dataset. Over several iterations of this procedure, the amount of intervention decreases and the policy performance improves.

Inverse reinforcement learning

Inverse reinforcement learning (IRL) (97) is another approach to IL that aims to deduce the underlying reward function that the expert surgeon is assumed to be optimizing. This approach offers several advantages: It provides insight into the reasons behind expert behavior, it can accommodate suboptimal expert demonstrations, and it allows for differences in robot–surgeon dynamics or capabilities. IRL typically assumes a parameterized reward function, often as a linear combination of nonlinear features.

IRL faces several challenges in surgical robotics. Multiple reward functions can often explain the same set of expert demonstrations, leading to ambiguity. IRL methods often operate under the assumption that the expert’s behavior is optimal, which can be problematic when human surgeons exhibit suboptimal or inconsistent behavior. In addition, IRL methods need to be scalable to complex surgical environments and capable of generalizing from limited data.

Imitation learning in surgery

Thus far, IL approaches with surgical robots (91) are generally performed either in simulation or on tabletop tasks, such as fundamentals of laparoscopic surgery (FLS) tasks. These environments provide a controlled setting that allows for the detailed capture and analysis of model behaviors, which are essential for model development (Fig. 4). Below, we discuss examples of relevant prior works.

Simulation

In simulation, IL methods have been used to enable surgical robots to refine their skills in a safe and controlled environment. Some works focus on IL on realistic tissue simulations, such as LapGym (74), which includes a gallbladder, abdominal cavity, and tissue dissection model. LapGym (74, 93), Orbit-Surgical (71), and SuRoL (76, 98, 99) have tabletop and FLS skill simulations for surgically relevant tasks such as peg pick-and-place and rope threading. The primary challenge with these methods is that they face the same sim-to-real challenge as the RL simulated solution.

Tabletop tasks

Outside of simulation, tabletop and FLS tasks are considered the first step toward teaching robots the necessary skills required during surgical procedures and have thus far received the most attention in IL. These tasks typically involve simpler operations, such as cutting, suturing, or manipulating tissues, which are fundamental to many surgical procedures.

Tanwani *et al.* (100) learned an embedding feature space from video frames by minimizing a metric learning loss (i.e., images from similar actions are clustered while pushed away from random images), which was used to perform end-effector pose imitation. In work from Kawaharazuka *et al.* (101), Panda robot arms with laparoscopic attachments were trained to perform peg transfer using monocular images. The Surgical Robot Transformer (SRT) (102) solved three challenging surgical tasks, including knot tying, tissue manipulation, and suture pick-up. Wrist cameras were used to improve performance, and a relative action representation was used to improve kinematic errors on the dVRK. Motion primitive diffusion was introduced by Scheikl *et al.* (93) to solve a tissue deformation problem. This method adapts motion primitives to existing diffusion-based techniques for gentle manipulation of deformable objects. Kim *et al.* (103) demonstrated reliable navigation to desired target points in retinal surgery by combining optimal control and imitation learning (104).

The primary challenge for these techniques is acquiring sufficient data to train an imitation policy. At the time of writing, kinematic data are not easily acquirable from existing surgical robots because of company regulations. Existing systems that are able to collect kinematic data (e.g., the dVRK or adapted industrial robots) are not approved for use in humans; thus, it is challenging to acquire sufficient expert surgical data. One possible direction for future work is to develop an accurate kinematics estimator from video frames, as has been done in prior works (105, 106), that could be used to collect kinematic data from real surgeries. Although IL leverages high-quality expert demonstration data to replicate sophisticated surgical behaviors, its effectiveness is limited by the availability and consistency of such data; furthermore, the risk of replicating suboptimal human actions necessitates careful curation and validation of demonstration datasets.

Combining imitation learning and reinforcement learning

RL and IL can be combined in several ways to leverage expert demonstrations for improved sample efficiency, exploration, and

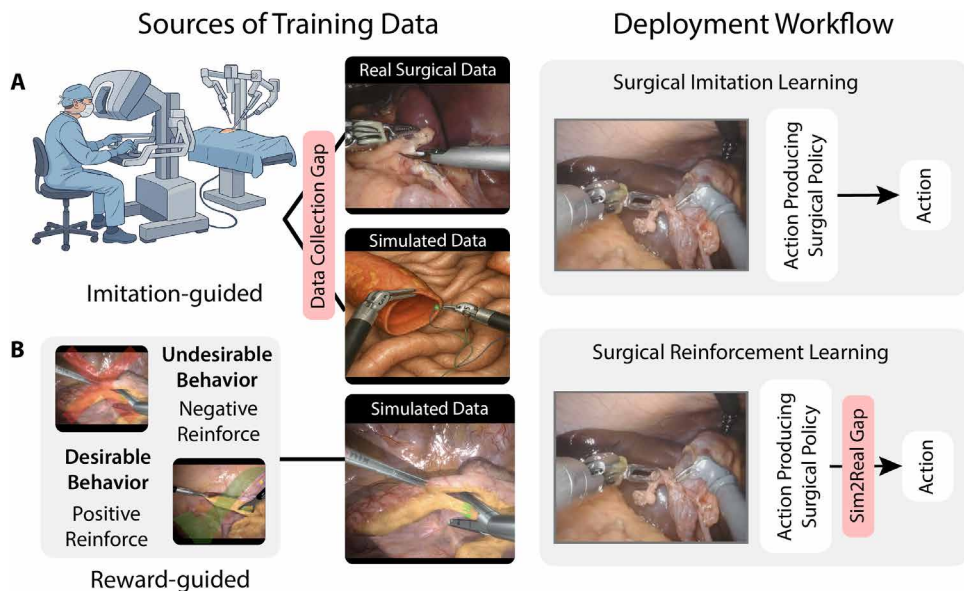


Fig. 4. Overview training data and deployment workflows for RL and IL. (A) Imitation relies on expert demonstrations, whereas (B) RL relies on randomly sampling behaviors and measuring the quality. IL suffers from the data collection gap, whereas RL suffers from the sim-to-real gap.

performance. Early methods such as (107–109) used demonstration data to augment standard RL algorithms. For instance, deep Q-learning from demonstrations (107) and deep deterministic policy gradient from demonstrations (108) bootstrap Q-learning and continuous control agents, respectively, from limited demonstration trajectories. Demonstration-augmented policy gradient (DAPG) (109) similarly combines BC with on-policy optimization, showing performance gains on complex robotic tasks.

Offline RL approaches also benefit from combining IL-inspired objectives. Conservative Q-learning (CQL) (110) avoids overestimating out-of-distribution actions by penalizing high Q-values on unseen state-action pairs, whereas TD3 + BC (111) explicitly adds a BC term to stabilize offline training. Related works like soft-Q imitation learning (SQIL) (112) and IQ-Learn (113) reframe imitation as sparse-reward Q-learning, effectively prioritizing expert transitions in offline datasets. Beyond purely algorithmic developments, the application of RL and IL in medical robotics has been limited, except one example by Keller *et al.* (114). In this work, the authors automated deep anterior lamellar keratoplasty on *ex vivo* human corneas using RL from demonstration (RLfD). Their findings demonstrated that the final trained policy not only achieved higher precision than that of surgical fellows but also improved consistency and accuracy in needle insertion depth, highlighting the potential of RLfD to improve surgical outcomes.

Case studies

We focus our attention on two representative works that explored promising strategies in IL (102) and RL (80). SRT (102), as shown in Fig. 5, falls in the category of IL that demonstrated complex surgical behavior in tabletop tasks using real-world data. Specifically, SRT demonstrated suture knot tying, needle manipulation, and soft-tissue manipulation using expert demonstration data and the dVRK system. Beyond tabletop settings, SRT showed zero-shot generalization of the trained models in unseen scenarios, such as knot tying on porcine tissues and needle manipulation on chicken tissue, although the models were trained using silicone suture pads only. SRT also addressed a fundamental issue, which involves dealing with the notorious inaccurate forward kinematics of the dVRK system. Specifically, the dVRK joint measurements are prone to errors because of a lack of precise motor encoders, hysteresis, and overall slack and flexibility in its mechanism. Given that the kinematics accuracy is low, the pool of surgical data collected using da Vinci systems is inevitably inconsistent, posing difficult challenges for robot learning. SRT showed that using such inconsistent kinematic data led to high failure rates, and the learned policies were not usable across all tasks. To alleviate this issue, SRT introduced tool-centric and hybrid-relative action representations, which enabled direct use of the inconsistent kinematic data, without performing further calibration or corrective steps to train a feasible policy. This work demonstrated promise for using large-scale approximate kinematic data for learning complex manipulation behaviors, which can be useful for leveraging the large repository of surgeon data to effectively build large-scale models for autonomous surgery.

Another promising work by Scheikl *et al.* (80), as shown in Fig. 5, falls in the category of RL while overcoming the sim-to-real gap to transfer policies trained in simulation to the real world. Specifically, (80) presented a pipeline that leveraged pixel-level domain adaptation through contrastive generative adversarial networks (GANs) to train visuomotor policies in simulation, which was successfully deployed in the real world on a dVRK without retraining. The core idea was to train

a model in simulation using translated images that looked like real-world images so that the sim-to-real gap was reduced when deployed in the real world. The study focused on the tissue retraction task, a critical step during tissue dissection. The proposed approach demonstrated a 50% success rate using raw RGB images, which is a promising result for sim-to-real strategies. This work not only demonstrated the potential for using RL in complex surgical tasks but also set a precedent for efficient transfer of skills from sim-to-real in robotic surgery, making RL a more viable tool for automating surgical tasks.

Comparison of reinforcement learning and imitation learning

RL and IL each present distinct advantages and limitations for increasing robotic autonomy in surgery, as illustrated in Fig. 3. RL relies on a simulator, but because simulating soft tissue deformation is challenging, these environments can lack realism, leading to difficulties in translating robot behaviors to real patients. IL relies on demonstrations by expert surgeons, but if the dataset lacks variety or contains suboptimal demonstrations, the learned policy will fail to generalize or will replicate undesirable behaviors.

Safety considerations also differ. RL leverages random exploration, which is infeasible in live surgical scenarios. Conversely, IL reproduces known expert trajectories, making early deployment safer but potentially overlooking rare complications not present in the original dataset. For adaptability, RL can discover strategies that deviate from standard practice, whereas IL may lack the capacity to adapt to out-of-distribution events.

Despite these differences, it is possible to combine RL and IL into a hybrid framework, merging the safety and stability of IL with the exploratory power of RL. One promising approach is to initialize the robot's policy by imitating expert demonstrations and then refine this policy through reinforcement.

We highlight representative performance metrics from prior IL and RL studies, primarily in laparoscopic settings. Kim *et al.* (86) used IL with a few hundred demonstrations per task and reported a 100% success rate in a needle pickup task and 90% in a knot tying task. In (81), an RL approach using proximal policy optimization (115) achieved 100% success in reaching tasks and 90% in rigid object manipulation. Similarly, Scheikl *et al.* (93) reported 78.6% success in rope manipulation and 99% in general object manipulation using diffusion models. These results suggest that, with sufficient data and in relatively simple tabletop environments, learned policies can consistently achieve 80 to 90% success rates in nontrivial manipulation tasks.

Control strategies

Cable-driven surgical robots are susceptible to hysteresis and backlash, which cause non-negligible kinematic errors (116). Although hand-eye calibration can compensate, it often requires repeated, impractical recalibration for different tools in time-sensitive settings (117–119). One strategy is online calibration; Li *et al.* (120) used a particle filter fusing kinematic priors with visual features to estimate tool pose in real time, even in deformable environments, while accounting for uncertainty. A simpler alternative avoids absolute kinematics altogether. In (86), policies used relative motion commands, enabling robust control where the robot continuously refined its motion with visual feedback instead of seeking a fixed, potentially inaccurate spatial target. For instance, a relative-motion policy can iteratively adjust a needle grasp, whereas an absolute policy might repeatedly fail because of kinematic drift.

Sensors used for autonomous surgical AI

Sensor technology is an important component of autonomous surgical systems. High-resolution imaging sensors—most commonly

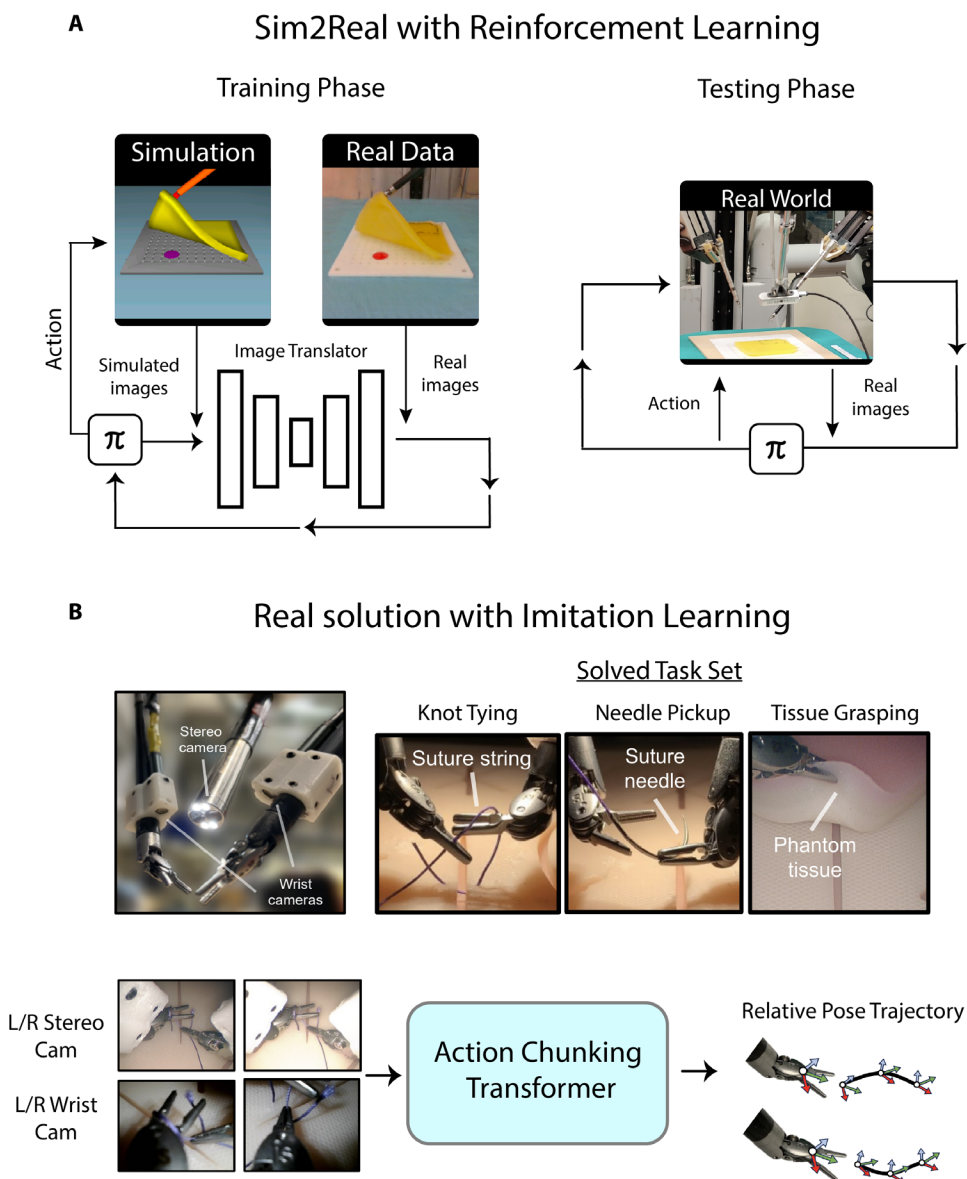


Fig. 5. Examples of two surgical systems that were successfully deployed on hardware. (A) Policy that was trained in simulation and used an image translator to synchronize simulation frames with real data. **(B)** Policy that used IL via an action chunking transformer to mimic surgical trajectories.

dual-camera stereoscopic endoscopes (86, 90)—provide the raw RGB data and coarse disparity cues that current learning algorithms rely on. In addition to vision, a variety of other sensors—including force-torque sensors (121) and tactile sensors (122, 123)—provide real-time feedback on tool-tissue interactions. The integration of these multimodal sensor streams enables precise depth perception, improved situational awareness, and adaptive control in dynamic surgical environments.

Nonetheless, today's vision and depth-sensing stack remains the rate-limiting step for reliable 3D understanding in vivo. Stereo endoscopes produce noisy disparity maps beyond a few centimeters and fail in specular or blood-covered regions. Time-of-flight and structured-light probes suffer from multipath interference inside fluid-filled cavities, whereas photometric methods are confounded by smoke,

glare, and tissue translucency. Existing reconstructions therefore drift, exhibit zipper artifacts, or collapse entirely when the scope pivots or the scene deforms. These shortcomings motivate algorithms that jointly reason over geometry, appearance, and instrument motion—for example, learning-based monocular depth estimation, differentiable simultaneous localization and mapping (SLAM) that fuses frame-to-frame geometry with kinematic priors, and neural implicit surfaces that update in real time. Pioneering systems such as STAR demonstrated submillimeter 3D point-cloud overlays yet required a rigidly mounted scope and precalibrated lighting; next-generation methods must generalize across freehand endoscope motion, variable illumination, and occlusions.

Despite these advances, reliable perception in a live operating room remains notoriously difficult. Blood splatter, tissue debris, and smoke plumes can occlude critical anatomy, whereas irrigation and insufflation introduce dynamic lighting changes and airborne particulates that confuse vision pipelines (124–126). State-of-the-art systems therefore fuse physics-based data augmentation, domain-randomization techniques that vary lighting and optical properties during training, and multispectral or polarization imaging to increase resilience (126–128). Learning-based artifact detection modules can identify active bleeding or smoke and trigger adaptive exposure, suction, or complementary sensing to preserve situational awareness (129, 130). Recent demonstrations of autonomous blood-clearing suction and robust sim-to-real transfer with aggressive camera lighting randomization illustrate practical pathways to overcoming these perception challenges (54).

THE FUTURE OF AI IN ROBOTIC SURGERY

Now, the primary approach in robot learning has been to try and solve task-specific goals that address individual challenges, such as, in surgery, autonomous suturing or needle pick-up. Recent developments, however, have trended toward high-capacity models that showed major improvements in performance when trained on extensive, diverse, and nonspecific imitation datasets (131–135). These models excel in adapting to new environments, a capability that grows with increases in data volume and model complexity (133).

This approach differs from what has been done with surgical robot systems, which depend on data-driven learning and have progressed more slowly compared with other areas in robotics. This delay can be attributed to a few key issues: the scarcity of large,

open-source datasets for training, challenges in modeling the soft-body deformations encountered during surgeries, and the increased risk of patient injury during clinical trials, which calls for stricter safety protocols (60). Although small task-level imitation policies with open datasets have been achieved (136, 137), there does not exist a larger repository of data to date. As suggested by new advances in robotics, a path toward improving the autonomy of surgical robots could be through the development of a large, multimodal, VLA model specifically for surgery.

In this section, we outline an approach toward building an “AI robotic surgeon” (Fig. 6) by first building a vision-language model for surgery trained on medical text and image content all the way from premedical studies through surgical residency. This model could then be trained using IL on kinematic data from surgical skills training, animal surgeries, and full-human surgeries. Last, this model could continually improve with experience. Below, we describe these ideas in depth.

Building a foundation for language in surgery

VLA models are typically built using language models as the foundation for further training (65, 138). Common foundations for language models include Llama 2 (139), Vicuna (140), and PaLM-E (141). These models are trained to imitate text from trillions of “tokens” (chunks of words), which represent human-generated natural language through a process called next-token prediction (142). This trained model is then often fine-tuned on instruction datasets (e.g., question-answering) in a process called instruct tuning (143–145). Because most VLAs are just performing household tasks involving common objects, now, the language models serving as the foundation of VLAs do not present issues for learning the task. However, building a VLA that can perform surgical operations introduces unique challenges.

The likely first step toward building an AI-based robotic surgeon would be training a language model on the same information that surgeons must understand before entering practice, such as a comprehensive medical curriculum from premedical training all the way through passing surgical board exams (Fig. 6). Such a model would have an understanding of both medical and surgical terms that would be important for understanding and performing surgery. Currently existing surgical language models have thus far focused on task-specific optimizations and lack a comprehensive understanding of broader concepts in biology, medicine, and even academic surgery (such as textbook question-answering) (146, 147).

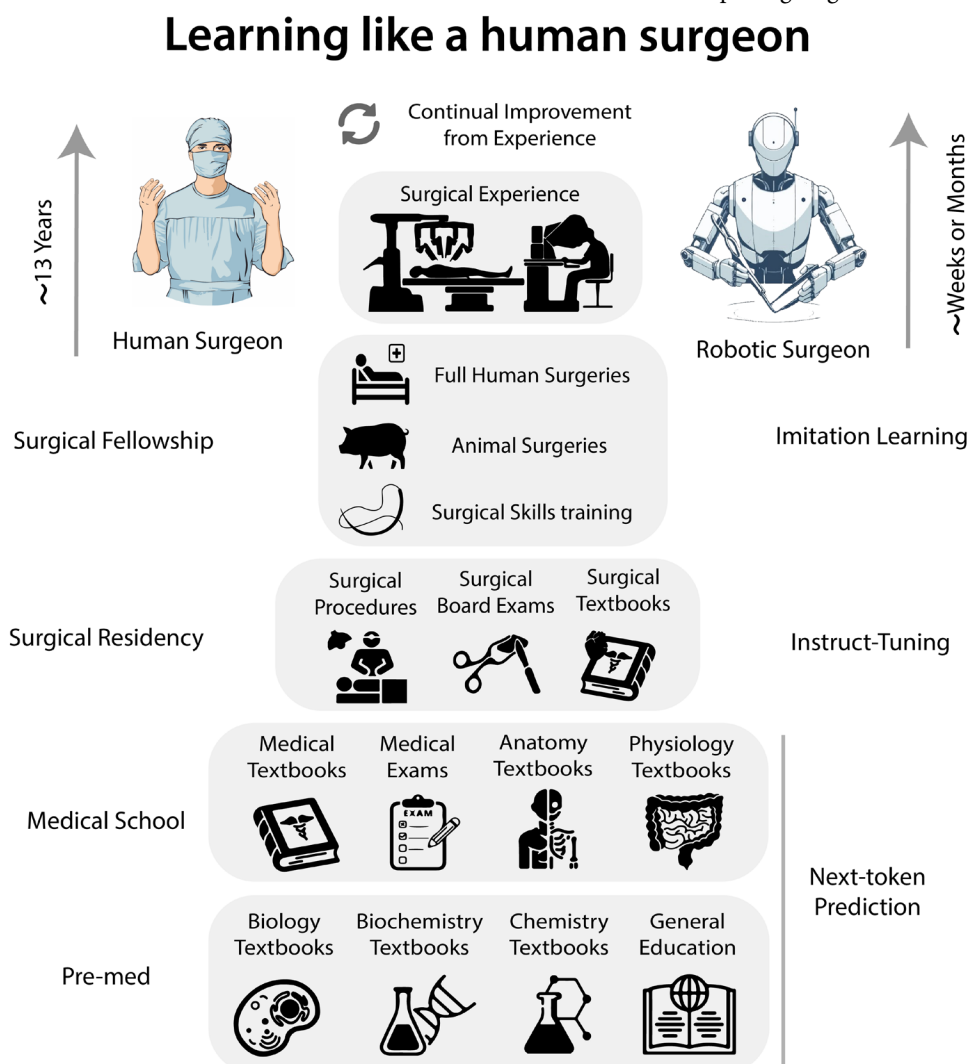


Fig. 6. Outline of the AI robotic surgeon. A language model is trained via next-token prediction and instruct tuning on the same information as surgeons all the way up through surgical residency. A VLA model is then trained via IL on surgical skills (e.g., suture pads), animal surgeries, and full human surgeries. Last, the model continually improves through experience using RL.

VLA models in robotics

VLAs represent a substantial shift from traditional task-specific robotics, integrating visual perception, language understanding, and action generation into a unified framework for more flexible systems (131–135). Typically based on large-scale transformers, VLAs are trained on diverse datasets of robot demonstrations with language instructions. Their key innovation is processing multimodal inputs—like images and text commands—to directly output low-level control actions in an end-to-end manner.

VLAs show impressive generalization to tasks not seen during training. With a large enough dataset, the model learns general correlations between language, vision, and action spaces, avoiding overfitting. For example, a robot can interpret an instruction like “find me a drink that will energize me” and retrieve an energy drink without explicitly training on that particular command (133).

VLAs in robotic surgery

Now, there do not exist any general-purpose VLAs in robotic surgery. However, the potential of VLA models in surgical robotics is particularly promising for a number of reasons (60). There are

already thousands of experts collecting teleoperated surgery data that could be used for imitation around the world every single day. In addition, surgical robots do not face the same limitations as mobile robots and readily support intensive onboard computers, which increases the potential capacity of VLA model size. By training on large datasets of surgical demonstrations paired with expert annotations, these models could potentially learn to interpret complex surgical scenarios and generate appropriate actions beyond human capability.

However, challenges remain in applying VLAs to surgical robotics. These include the need for large, high-quality datasets of annotated surgical procedures, ensuring safety and reliability in high-stakes medical environments, and addressing the interpretability of model decisions.

Measurable safety

One of the primary capabilities that would need to be embedded in a surgical robot is the ability to recognize its own limitations during a surgical procedure and appropriately divert control back to a human surgeon. This need arises particularly in scenarios that are not covered by the robot's training, necessitating intervention by a human. Unlike in conventional engineering, where potential problems are exhaustively tested by simulating every conceivable scenario, the unpredictable nature of surgical procedures makes such comprehensive preparation impractical. Consequently, surgical robots must be designed with the capability to handle previously unencountered situations. Robots, particularly in their early developmental stages, cannot be expected to adapt at a level comparable to that of human surgeons.

To mitigate these issues, one potential approach is to use the reinforcement-based technique called CQL (110, 148), a method that trains the robot to steer clear of situations not present in their training data. CQL focuses on developing a conservative estimate of value functions to avoid overestimation, which is common in offline deep RL. This method has been successfully applied in some robot transformer architectures, which was shown in a demonstration where the robot asked for clarification with ambiguous instructions. An alternative method, conformal prediction (149, 150), offers another solution by providing a range of probable outcomes for each decision, thus informing users of the associated uncertainty levels. This method splits incoming data into a training set and a calibration set, with the latter fine-tuning the confidence levels for predictions. Both CQL and conformal prediction, if integrated, could be crucial in enabling surgical robots to avoid risky scenarios and defer to surgeons when necessary.

Adversarial robustness is critical because of vulnerabilities in complex AI models, particularly in perception (151). Adversarial attacks, via subtle perturbations to sensor inputs, can cause misclassifications or faulty predictions by the AI model, leading to erroneous robotic actions (152, 153). Mitigating these threats requires robust model training techniques, secure data pipelines, and run-time monitoring (154). The necessary hardware and software architectures must work in symbiosis to ensure both high-performance computation and verifiable safety. This involves integrating specialized hardware like AI accelerators and secure processing units with safety-certified software stacks, which may use real-time operating systems, hypervisors for a mixed-criticality approach, and securely containerized environments for deploying AI models (155). Formal verification methods are also important for demonstrating system integrity and reliability.

The role of regulation in AI-based surgery

Understanding and developing a framework for regulatory approval for AI in surgery will be essential to ensure the proper testing, evaluation, and compliance of an AI system to prioritize and maintain patient safety and efficacy in the operating room. The FDA has seen an exponential increase in the number of regulatory submissions related to AI medical devices in recent years, rising from only six approved devices in 2015 to more than 200 devices in 2023 alone (156). This trend has been driven by more connected devices, more commercial investment in technology, and a growing familiarity of how software can be regulated as a medical device. Today, the FDA has approved more than 950 AI-based products, but most of these products are related to signal and image processing, such as Siemens Healthineers' AI-Rad Companion (157), which provides qualitative and quantitative information of clinical images, and Digital Diagnostics' LumineticsCore (158), which analyzes images to detect diabetic retinopathy.

As AI becomes a decision-making platform in surgery, new regulatory oversight is needed. In January 2025, the FDA released draft guidance for AI-enabled medical devices, addressing transparency, bias, and product design (159). The FDA is adopting a total product life cycle approach, meaning that AI devices must meet regulatory standards throughout their entire life span, not just at initial approval. To manage evolving AI models, the guidance introduces predetermined change control plans, which allow companies to predefine software updates without seeking new clearance for each change, streamlining the regulatory process.

The guidance also mandates transparency, requiring clear explanations of how AI systems make decisions to build trust among clinicians and patients. Furthermore, the FDA emphasizes that models must be tested across diverse populations to avoid bias and ensure equitable outcomes. The adoption of these regulations will require input from the medical device community, and feedback from key opinion leaders is paramount to ensure that AI regulation is managed efficiently by both industry and government to guarantee safe and effective AI-driven systems for the future.

Barriers to entry, clinical deployment, and integration challenges

Limitations in today's regulatory process are a rate-limiting step in adopting AI surgical robots. New FDA guidance mandates a total life cycle policy, requiring perpetual oversight, validation, and post-market surveillance. To successfully market AI devices, manufacturers must engage the FDA early, develop well-defined predetermined change control plans that outline software validation, and ensure robust postmarket surveillance protocols.

To limit market entry barriers beyond regulatory clearance, companies will likely introduce AI systems in a graduated fashion. For instance, Intuitive Surgical released AI software that operates in the background, analyzing system data to create objective insights for surgeons. Because the AI only analyzes data and leaves decision-making to the surgeon, it is easier to regulate and deploy. This trend is already unfolding with Moon Surgical's Maestro system, an AI-based laparoscope holder that received the first FDA clearance for an intraoperative AI application. Because the task does not interfere with a surgeon's workflow, it reduces barriers to clinical adoption. If successful, other companies will likely follow this approach, slowly building a new infrastructure where AI-based tasks in the operating room become common.

Beyond regulatory and workflow considerations, the clinical translation of autonomous surgical robots faces substantial institutional hurdles, including determining potential liability. Hospitals must invest in infrastructure upgrades, including robust data management systems, high-performance computing, and potentially redesigned operating rooms. Integrating these systems into established workflows requires careful planning; adaptation of team roles; and comprehensive training for surgeons, nurses, and technicians. Furthermore, system maintenance, software updates, specialized sterilization protocols, and overall cost-effectiveness will be critical factors influencing widespread adoption and accessibility.

Acquiring surgical datasets

Surgical video data are extensively available from various procedures, including cataracts (160, 161), neurosurgery (162), and cholecystectomy (163–165). There are also data accessible from general manipulation skills like peg transferring with laparoscopic tools (137, 166, 167). In addition, public platforms like YouTube host numerous surgical videos, with one curated dataset containing 2000 open-surgery demonstrations (168). Despite all of these open data, acquiring data for large-scale surgery projects remains challenging, primarily because of patient privacy concerns and the typically small sample sizes in medical studies.

Recent collaborative efforts in general robotics have shown promise in overcoming these data limitations. A project involving 21 institutions assembled a dataset to train a large-scale RT model capable of controlling multiple robots across diverse tasks (169). This work demonstrates the potential for high-capacity models to improve with increased data, even beyond their expected applications. However, replicating such large-scale data collection in surgical settings poses challenges. Successful large-scale collaborations in medicine suggest that realizing a surgical VLA would require extensive data sharing and open collaboration among universities, industry, and hospitals.

CONCLUSION

AI and robotics are enabling surgical autonomy, with the potential to improve outcomes and expand access to care. Surgical robots with built-in autonomy present a promising solution to current challenges, offering the potential to improve efficiency, safety, and consistency in surgical procedures.

We divide current research focuses into two primary AI approaches: simulation-based learning using RL techniques and imitation-based learning that aims to replicate expert surgeon behaviors. Although both approaches show promise, they face considerable challenges. Simulation-based methods struggle with the sim-to-real gap, where models trained in simplified virtual environments may not translate effectively to real-world surgical scenarios. Imitation-based learning, although potentially more adaptable, requires extensive high-quality datasets of expert demonstrations and may inadvertently replicate sub-optimal behaviors.

A path toward improving the autonomy of surgical robots could be through the development of large, multimodal VLAs specifically for surgery. These models integrate visual perception, natural language understanding, and action generation into a unified framework. Challenges remain in applying VLAs to surgical robotics, including the need for large, high-quality datasets; ensuring safety and reliability in high-stakes medical environments; and addressing the interpretability of model decisions. A primary capability for a

surgical robot is the ability to recognize its own limitations during a procedure and appropriately divert control back to a human surgeon. As the field moves forward, it will be essential to maintain a balance between technological innovation and patient safety, ensuring that the development of autonomous surgical robots ultimately leads to improved patient outcomes and increased access to high-quality surgical care worldwide.

REFERENCES AND NOTES

- R. L. Siegel, N. S. Wagle, A. Cercek, R. A. Smith, A. Jemal, Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 233–254 (2023).
- H. J. Lujan, G. Plasencia, M. Jacobs, M. Viamonte, R. F. Hartmann, Long-term survival after laparoscopic colon resection for cancer. *Dis. Colon Rectum* **45**, 491–501 (2002).
- A. Ehrampoosh, B. Shirinzadeh, J. Pinski, J. Smith, R. Moshinsky, Y. Zhong, A force-feedback methodology for teleoperated suturing task in robotic-assisted minimally invasive surgery. *Sensors* **22**, 7829 (2022).
- J. E. Knudsen, U. Ghaffar, R. Ma, A. J. Hung, Clinical applications of artificial intelligence in robotic surgery. *J. Robot. Surg.* **18**, 102 (2024).
- G.-Z. Yang, Digital architecture and robotic construction. *Sci. Robot.* **2**, ean3673 (2017).
- P. E. Dupont, B. J. Nelson, M. Goldfarb, B. Hannaford, A. Mencias, M. K. O'Malley, N. Simaan, P. Valdastrì, G. Z. Yang, A decade retrospective of medical robotics research from 2010 to 2020. *Sci. Robot.* **6**, eabi8017 (2021).
- J. J. Rasouli, J. Shao, S. Neifert, W. N. Gibbs, G. Habboub, M. P. Steinmetz, E. Benzel, T. E. Mroz, Artificial intelligence and robotics in spine surgery. *Global Spine J.* **11**, 556–564 (2021).
- A. Moglia, K. Georgiou, E. Georgiou, R. M. Satava, A. Cuschieri, A systematic review on artificial intelligence in robot-assisted surgery. *Int. J. Surg.* **95**, 106151 (2021).
- K. Denecke, C. R. Baudoin, A review of artificial intelligence and robotics in transformed health ecosystems. *Front. Med.* **9**, 795957 (2022).
- T. Haidegger, Autonomy for surgical robots: Concepts and paradigms. *IEEE Trans. Med. Robot. Bionics* **1**, 65–76 (2019).
- S. Panesar, Y. Cagle, D. Chander, J. Morey, J. Fernandez-Miranda, M. Klot, Artificial intelligence and the future of surgical robotics. *Ann. Surg.* **270**, 223–226 (2019).
- I. Andras, E. Mazzone, F. W. B. van Leeuwen, G. de Naeyer, M. N. van Oosterom, S. Beato, T. Buckle, S. O'Sullivan, P. J. van Leeuwen, A. Beulens, N. Crisan, F. D'Hondt, P. Schattman, H. van der Poel, P. Dell'Oglio, A. Mottrie, Artificial intelligence and robotics: A combination that is changing the operating room. *World J. Urol.* **38**, 2359–2366 (2020).
- A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, P. Valdastrì, Autonomy in surgical robotics. *Annu. Rev. Control Robot. Auton. Syst.* **4**, 651–679 (2021).
- B. Kuipers, E. A. Feigenbaum, P. E. Hart, N. J. Nilsson, Shakey: From conception to history. *AI Mag.* **38**, 88–103 (2017).
- R. D. Howe, Y. Matsuoka, Robotics for surgery. *Annu. Rev. Biomed. Eng.* **1**, 211–240 (1999).
- A. Gawande, Two hundred years of surgery. *N. Engl. J. Med.* **366**, 1716–1723 (2012).
- A. Buia, F. Stockhausen, E. Hanisch, Laparoscopic surgery: A qualified systematic review. *World J. Methodol.* **5**, 238–254 (2015).
- Y. Zhou, H. Ren, M. Q.-H. Meng, Z. T. H. Tse, H. Yu, Robotics in natural orifice transluminal endoscopic surgery. *J. Mech. Med. Biol.* **13**, 1350044 (2013).
- S. Atallah, B. Martin-Perez, D. Keller, J. Burke, L. Hunter, Natural-orifice transluminal endoscopic surgery. *J. Br. Surg.* **102**, e73 (2015).
- D. Ramirez, M. J. Maurice, J. H. Kaouk, Robotic single-port surgery: Paving the way for the future. *Urology* **95**, 5–10 (2016).
- R. J. Nelson, J. S. S. Chavali, N. Yerram, P. Babbar, J. H. Kaouk, Current status of robotic single-port surgery. *Urol. Ann.* **9**, 217–222 (2017).
- C. P. Childers, M. Maggard-Gibbons, Estimation of the acquisition and operating costs for robotic surgery. *JAMA* **320**, 835–836 (2018).
- Intuitive Surgical, Intuitive Annual Report 2024; <http://investor.intuitivesurgical.com/static-files/500ff989-ad91-4b32-a59e-f94a34d75997>.
- M. Kawka, Y. Fong, T. M. H. Gall, Laparoscopic versus robotic abdominal and pelvic surgery: A systematic review of randomised controlled trials. *Surg. Endosc.* **37**, 6672–6681 (2023).
- A. P. Ng, Y. Sanaïha, S. S. Bakhtiyar, S. Ebrahimian, C. Branche, P. Benharash, National analysis of cost disparities in robotic-assisted versus laparoscopic abdominal operations. *Surgery* **173**, 1340–1345 (2023).
- I. Kassite, T. Bejan-Angoulvant, H. Lardy, A. Binet, A systematic review of the learning curve in robotic surgery: Range and heterogeneity. *Surg. Endosc.* **33**, 353–365 (2019).
- S. K. Pasquali, D. Thibault, S. M. O'Brien, J. P. Jacobs, J. W. Gaynor, J. C. Romano, M. Gaies, K. D. Hill, M. L. Jacobs, D. M. Shahian, C. L. Backer, J. E. Mayer, National variation in congenital heart surgery outcomes. *Circulation* **142**, 1351–1360 (2020).
- T. Xu, M. A. Makary, E. al Kazzi, M. Zhou, T. M. Pawlik, S. M. Hutfless, Surgeon-level variation in postoperative complications. *J. Gastrointest. Surg.* **20**, 1393–1399 (2016).

29. G. F. Sheldon, T. C. Ricketts, A. Charles, J. King, E. P. Fraher, A. Meyer, The global health workforce shortage: Role of surgeons and other providers. *Adv. Surg.* **42**, 63–85 (2008).
30. Y.-Z. Zhou, C.-Q. Wang, M.-H. Zhao, Z.-Y. Li, D. Chen, A.-L. Lian, Y. Ma, Surgical smoke: A hidden killer in the operating room. *Asian J. Surg.* **46**, 3447–3454 (2023).
31. W. Truscott, Impact of microscopic foreign debris on post-surgical complications. *Surg. Technol. Int.* **12**, 34–46 (2004).
32. T. G. Manning, M. Perera, D. Christidis, N. Kinnear, S. McGrath, R. O’Beirne, P. Zotov, D. Bolton, N. Lawrentschuk, Visual occlusion during minimally invasive surgery: A contemporary review of methods to reduce laparoscopic and robotic lens fogging and other sources of optical loss. *J. Endourol.* **31**, 327–333 (2017).
33. Y. Ou, A. Soleymani, X. Li, M. Tavakoli, Autonomous blood suction for robot-assisted surgery: A sim-to-real reinforcement learning approach. *IEEE Robot. Autom. Lett.* **9**, 7246–7253 (2024).
34. D. B. Kaber, M. R. Endsley, The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theor. Issues Ergon. Sci.* **5**, 113–153 (2004).
35. Z. Liao, X.-D. Duan, L. Xin, L.-M. Bo, X.-H. Wang, G.-H. Xiao, L.-H. Hu, S.-L. Zhuang, Z.-S. Li, Feasibility and safety of magnetic-controlled capsule endoscopy system in examination of human stomach: A pilot study in healthy volunteers. *J. Interv. Gastroenterol.* **2**, 155–160 (2012).
36. S. DiMaio, M. Hanuschik, U. Kreaden, “The da Vinci surgical system” in *Surgical Robotics: Systems Applications and Visions*, J. Rosen, B. Hannaford, R. M. Satava, Eds. (Springer, 2011), pp. 199–217.
37. J. C.-Y. Ngu, C. B.-S. Tsang, D. C.-S. Koh, The da Vinci Xi: A review of its capabilities, versatility, and potential role in robotic colorectal surgery. *Robot. Surg.* **4**, 77–85 (2017).
38. F. Celotto, N. Ramacciotti, A. Mangano, G. Danieli, F. Pinto, P. Lopez, A. Ducas, J. Cassiani, L. Morelli, G. Spolverato, F. M. Bianco, Da Vinci single-port robotic system current application and future perspective in general surgery: A scoping review. *Surg. Endosc.* **38**, 4814–4830 (2024).
39. N. E. Samalavicius, V. Janusonis, R. Siaulyis, M. Jasėnas, O. Deduchovas, R. Venckus, V. Ezerškiene, R. Paskėvičiute, G. Klimavičiute, Robotic surgery using Senhance® robotic platform: Single center experience with first 100 cases. *J. Robot. Surg.* **14**, 371–376 (2019).
40. G. Alshaiikh, R. M. Schols, J. A. Wolfs, R. Cau, T. J. van Mulken, “A dedicated robotic system for open (super-)microsurgery” in *Robotics in Plastic and Reconstructive Surgery*, J. C. Selber, Ed. (Springer, 2021), pp. 139–153.
41. C. Song, X. Ma, X. Xia, P. W. Y. Chiu, C. C. N. Chong, Z. Li, A robotic flexible endoscope with shared autonomy: A study of mockup cholecystectomy. *Surg. Endosc.* **34**, 2730–2741 (2020).
42. J. Gilbert, The EndoAssist™ robotic camera holder as an aid to the introduction of laparoscopic colorectal surgery. *Ann. R. Coll. Surg. Engl.* **91**, 389–393 (2009).
43. T. Bensignor, G. Morel, D. Reversat, D. Fuks, B. Gayet, Evaluation of the effect of a laparoscopic robotized needle holder on ergonomics and skills. *Surg. Endosc.* **30**, 446–454 (2016).
44. Z. Xu, T. D. Khokhlova, C. S. Cho, V. A. Khokhlova, Histotripsy: A method for mechanical tissue ablation with ultrasound. *Annu. Rev. Biomed. Eng.* **26**, 141–167 (2024).
45. L. O. Mair, X. Liu, B. Dandamudi, K. Jain, S. Chowdhury, J. Weed, Y. Diaz-Mercado, I. N. Weinberg, A. Krieger, MagnetoSuture: Tetherless manipulation of suture needles. *IEEE Trans. Med. Robot. Bionics* **2**, 206–215 (2020).
46. S. Yang, R. A. MacLachlan, C. N. Riviere, Manipulator design and operation of a six-degree-of-freedom handheld tremor-canceling microsurgical instrument. *IEEE/ASME Trans. Mechatron.* **20**, 761–772 (2015).
47. P. J. Wijsman, L. Molenaar, F. J. Voskens, C. D. van’t Hullenaar, I. A. Broeders, Image-based laparoscopic camera steering versus conventional steering: A comparison study. *J. Robot. Surg.* **16**, 1157–1163 (2022).
48. M. Roche, The MAKO robotic-arm knee arthroplasty system. *Arch. Orthop. Trauma Surg.* **141**, 2043–2047 (2021).
49. K. Nakao, B. Thavara, R. Tanaka, Y. Yamada, G. Joshi, K. Miyatani, T. Kawase, Y. Kato, Surgeon experience of the surgical safety with kinevo 900 in vascular neurosurgery: The initial experience. *Asian J. Neurosurg.* **15**, 464–467 (2020).
50. Y.-F. Cheng, H.-C. Chen, P.-C. Ke, W.-H. Hung, C.-Y. Cheng, C.-H. Lin, B.-Y. Wang, Image-guided video-assisted thoracoscopic surgery with Artis Pheno for pulmonary nodule resection. *J. Thorac. Dis.* **12**, 1342–1349 (2020).
51. J. Reisenauer, M. J. Simoff, M. A. Pritchett, D. E. Ost, A. Majid, C. Keyes, R. F. Casal, M. S. Parikh, J. Diaz-Mendoza, S. Fernandez-Bussy, E. E. Folch, Ion: Technology and techniques for shape-sensing robotic-assisted bronchoscopy. *Ann. Thorac. Surg.* **113**, 308–315 (2022).
52. A. Kuntz, M. Emerson, T. E. Ertop, I. Fried, M. Fu, J. Hoelscher, M. Rox, J. Akulian, E. A. Gillaspie, Y. Z. Lee, F. Maldonado, R. J. Webster III, R. Alterovitz, Autonomous medical needle steering in vivo. *Sci. Robot.* **8**, ead7614 (2023).
53. A. Murali, S. Sen, B. Kehoe, A. Garg, S. M. Farland, S. Patil, “Learning by observation for surgical subtasks: Multilateral cutting of 3D viscoelastic and 2D orthotropic tissue phantoms” in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2015), pp. 1202–1209.
54. F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, M. C. Yip, Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection. *IEEE Robot. Autom. Lett.* **6**, 1383–1390 (2021).
55. H. Saeidi, J. D. Opfermann, M. Kam, S. Wei, S. Leonard, M. H. Hsieh, J. U. Kang, A. Krieger, Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Sci. Robot.* **7**, eabj2908 (2022).
56. R. H. Taylor, B. D. Mittelstadt, H. A. Paul, W. Hanson, P. Kazanzides, J. F. Zuhars, B. Williamson, B. L. Musits, E. Glassman, W. L. Bargar, An image-directed robotic system for precise orthopaedic surgery. *IEEE Trans. Robot. Autom.* **10**, 261–275 (1994).
57. W. Kilby, J. R. Dooley, G. Kuduvalli, S. Sayeh, C. R. Maurer Jr., The CyberKnife® Robotic Radiosurgery System in 2010. *Technol. Cancer Res. Treat.* **9**, 433–452 (2010).
58. R. J. Harris, J. B. Mygatt, S. I. Harris, Systems and methods for autonomous intravenous needle insertion, US Patent 9,364,171 (2016).
59. R. Montés-Micó, A. Cerviño, T. Ferrer-Blasco, VisuMax® femtosecond laser for corneal refractive surgery. *Expert Rev. Ophthalmol.* **3**, 385–388 (2008).
60. S. Schmidgall, J. W. Kim, A. Kuntz, A. E. Ghazi, A. Krieger, General-purpose foundation models for increased autonomy in robot-assisted surgery. arXiv:2401.00678 [cs.RO] (2024).
61. A. Lee, T. S. Baker, J. B. Bederson, B. I. Rapoport, Levels of autonomy in FDA-cleared surgical robots: A systematic review. *NPJ Digit. Med.* **7**, 103 (2024).
62. A. Shademan, R. S. Decker, J. D. Opfermann, S. Leonard, A. Krieger, P. C. W. Kim, Supervised autonomous robotic soft tissue surgery. *Sci. Transl. Med.* **8**, 337ra64 (2016).
63. G. Fagogenis, M. Mencattelli, Z. Machaidze, B. Rosa, K. Price, F. Wu, V. Weixler, M. Saeed, J. E. Mayer, P. E. Dupont, Autonomous robotic intracardiac catheter navigation using haptic vision. *Sci. Robot.* **4**, eaaw1977 (2019).
64. P. A. York, R. Peña, D. Kent, R. J. Wood, Microbotic laser steering for minimally invasive surgery. *Sci. Robot.* **6**, eabd5476 (2021).
65. Y. Ma, Z. Song, Y. Zhuang, J. Hao, I. King, A survey on vision-language-action models for embodied AI. arXiv:2405.14093 [cs.RO] (2024).
66. R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction. *Robotica* **17**, 229–235 (1999).
67. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, ed. 2, 2018).
68. C. J. C. H. Watkins, P. Dayan, Q-learning. *Mach. Learn.* **8**, 279–292 (1992).
69. S. Schmidgall, A. Krieger, J. Eshraghian, Surgical gym: A high-performance GPU-based platform for reinforcement learning with surgical robots. arXiv:2310.04676 [cs.RO] (2023).
70. Y. Ou, S. Zargazadeh, P. Sedighi, M. Tavakoli, A realistic surgical simulator for non-rigid and contact-rich manipulation in surgeries with the da Vinci research kit. arXiv:2404.05888 [cs.RO] (2024).
71. Q. Yu, M. Moghani, K. Dharmarajan, V. Schorp, William Chung-Ho Panitch, J. Liu, K. Hari, H. Huang, M. Mittal, K. Goldberg, A. Garg, Orbit-surgical: An open-simulation framework for learning surgical augmented dexterity. arXiv:2404.16027 [cs.RO] (2024).
72. P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, S. P. DiMaio, “An open-source research kit for the da Vinci® surgical system” in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2014), pp. 6434–6439.
73. E. Tagliabue, A. Pore, D. Dall’Alba, M. Piccinelli, P. Fiorini, “UnityFlexML: Training reinforcement learning agents in a simulated surgical environment” in *I-RIM 2020 Conference Proceedings* (I-RIM, 2020), pp. 23–24.
74. P. M. Scheikl, B. Gyenes, R. Younis, C. Haas, G. Neumann, M. Wagner, F. Mathis-Ullrich, LapGym—An open source framework for reinforcement learning in robot-assisted laparoscopic surgery. arXiv:2302.09606 [cs.RO] (2023).
75. “PhysX physics engine”; <https://developer.nvidia.com/physx-sdk>.
76. J. Xu, B. Li, B. Lu, Y.-H. Liu, Q. Dou, P.-A. Heng, “SurRol: An open-source reinforcement learning centered and dVRK compatible platform for surgical robot learning” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2021), pp. 1821–1828.
77. W. Zhao, J. P. Queralt, T. Westerlund, “Sim-to-real transfer in deep reinforcement learning for robotics: A survey” in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (IEEE, 2020), pp. 737–744.
78. S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, K. Bousmalis, “Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2019), pp. 12619–12629.
79. J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2017), pp. 23–30.
80. P. M. Scheikl, E. Tagliabue, B. Gyenes, M. Wagner, D. Dall’Alba, P. Fiorini, F. Mathis-Ullrich, Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery. *IEEE Robot. Autom. Lett.* **8**, 560–567 (2023).

81. Q. Yu, M. Moghani, K. Dharmarajan, V. Schorp, W. Panitch, J. Liu, K. Hari, H. Huang, M. Mittal, K. Goldberg, A. Garg, "Orbit-Surgical: An open-simulation framework for learning surgical augmented dexterity" in *2024 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2024), pp. 15509–15516.
82. Y. Ou, M. Tavakoli, Sim-to-real surgical robot learning and autonomous planning for internal tissue points manipulation using reinforcement learning. *IEEE Robot. Autom. Lett.* **8**, 2502–2509 (2023).
83. F. Liu, Z. Li, Y. Han, J. Lu, F. Richter, M. C. Yip, "Real-to-sim registration of deformable soft tissue with position-based dynamics for surgical robot autonomy" in *2021 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2021), pp. 12328–12334.
84. J. Lu, F. Richter, M. C. Yip, Pose estimation for robot manipulators via keypoint optimization and sim-to-real transfer. *IEEE Robot. Autom. Lett.* **7**, 4622–4629 (2022).
85. M. Haiderbhai, R. Gondokaryono, T. Looi, J. M. Drake, L. A. Kahrs, "Robust sim2real transfer with the da Vinci research kit: A study on camera, lighting, and physics domain randomization" in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2022), pp. 3429–3435.
86. J. Kim, T. Zhao, S. Schmidgall, A. Deguet, M. Kobilarov, C. Finn, A. Krieger, "Surgical robot transformer (SRT): Imitation learning for surgical tasks" in *Proceedings of the 8th Conference on Robot Learning* (PMLR, 2024), pp. 130–134.
87. A. Hussein, M. M. Gaber, E. Elyan, C. Jayne, Imitation learning: A survey of learning methods. *ACM Comput. Surv.* **50**, 1–35 (2018).
88. D. A. Pomerleau, "ALVINN: An autonomous land vehicle in a neural network" in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed. (Morgan-Kaufmann, 1988), pp. 305–313.
89. J. Li, Y. Jin, Y. Chen, H. Yip, M. Scheppach, P. Chiu, Y. Yam, H. Meng, Q. Dou, "Imitation learning from expert video data for dissection trajectory prediction in endoscopic surgical procedure" in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2023), pp. 494–504.
90. B. Li, R. Wei, J. Xu, B. Lu, C. Yee, C. Ng, P. Heng, Q. Dou, Y. Liu, "3D perception based imitation learning under limited demonstration for laparoscope control in robotic surgery" in *2022 International Conference on Robotics and Automation (ICRA)* (IEEE, 2022), pp. 7664–7670.
91. S. Schmidgall, J. W. Kim, A. Krieger, Robots learning to imitate surgeons—challenges and possibilities. *Nat. Rev. Urol.* **21**, 451–452 (2024).
92. T. Z. Zhao, J. Tompson, D. Driess, P. Florence, K. Ghasemipour, C. Finn, A. Wahid, ALOHA unleashed: A simple recipe for robot dexterity. arXiv:2410.13126 [cs.RO] (2024).
93. P. M. Scheikl, N. Schreiber, C. Haas, N. Freymuth, G. Neumann, R. Lioutikov, F. Mathis-Ullrich, Movement primitive diffusion: Learning gentle robotic manipulation of deformable objects. *IEEE Robot. Autom. Lett.* **9**, 5338–5345 (2024).
94. S. Ross, G. Gordon, D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning" in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, M. Dudík, Eds., vol. 15 of *Proceedings of Machine Learning Research* (PMLR, 2011), pp. 627–635.
95. M. Kelly, C. Sidrane, K. Driggs-Campbell, M. J. Kochenderfer, "HG-Dagger: Interactive imitation learning with human experts" in *2019 International Conference on Robotics and Automation (ICRA)* (IEEE, 2019), pp. 8077–8083.
96. K. Menda, K. Driggs-Campbell, M. J. Kochenderfer, "EnsembleDagger: A Bayesian approach to safe imitation learning" in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2019), pp. 5041–5048.
97. A. Y. Ng, S. J. Russell, "Algorithms for inverse reinforcement learning" in *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning* (Morgan Kaufmann Publishers Inc., 2000), pp. 663–670.
98. T. Huang, K. Chen, B. Li, Y.-H. Liu, Q. Dou, "Demonstration-guided reinforcement learning with efficient exploration for task automation of surgical robot" in *2023 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2023), pp. 4640–4647.
99. Y. Long, W. Wei, T. Huang, Y. Wang, Q. Dou, Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning. *IEEE Robot. Autom. Lett.* **8**, 4441–4448 (2023).
100. A. K. Tanwani, P. Sermanet, A. Yan, R. Anand, M. Phielipp, K. Goldberg, "Motion2Vec: Semi-supervised representation learning from surgical videos" in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2020), pp. 2174–2181.
101. K. Kawaharazuka, K. Okada, M. Inaba, "Robotic constrained imitation learning for the peg transfer task in fundamentals of laparoscopic surgery" in *2024 IEEE International Conference on Robotics and Automation* (IEEE, 2024), pp. 606–612.
102. J. W. Kim, S. Schmidgall, A. Krieger, M. Kobilarov, "Learning a library of surgical manipulation skills for robotic surgery," presented at the workshop Bridging the Gap between Cognitive Science and Robot Learning in the Real World: Progresses and New Directions, Conference on Robot Learning, Atlanta, GA, 6 to 9 November 2023.
103. J. W. Kim, C. He, M. Urias, P. Gehlbach, G. D. Hager, I. Iordachita, M. Kobilarov, "Autonomously navigating a surgical tool inside the eye by learning from demonstration" in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2020), pp. 7351–7357.
104. J. W. Kim, P. Zhang, P. Gehlbach, I. Iordachita, M. Kobilarov, "Towards autonomous eye surgery by combining deep imitation learning with optimal control" in *Proceedings of the 2020 Conference on Robot Learning*, J. Kober, F. Ramos, C. Tomlin, Eds., vol. 155 of *Proceedings of Machine Learning Research* (PMLR, 2020), pp. 2347–2358.
105. J. Romero, D. Tzionas, M. J. Black, Embodied hands. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **36**, 1–17 (2017).
106. K. Shaw, S. Bahl, D. Pathak, "VideoDex: Learning dexterity from internet videos" in *Proceedings of the 6th Conference on Robot Learning*, K. Liu, D. Kulic, J. Ichnowski, Eds., vol. 205 of *Proceedings of Machine Learning Research* (PMLR, 2023), pp. 654–665.
107. T. Hester, M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, I. Osband, G. Dulac-Arnold, J. Agapiou, J. Leibo, A. Grusl, "Deep Q-learning from demonstrations" in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI, 2018), pp. 3223–3230.
108. M. Vecerik, T. Hester, J. Scholz, F. Wang, O. Pietquin, A. Barto, J. Hwang, B. Piot, N. Heess, T. Rothörl, T. Lampe, M. Riedmiller, Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards. arXiv:1707.08817 [cs.AI] (2018).
109. A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations" in *Robotics: Science and Systems (RSS)* (RSS Foundation, 2018), 10.15607/RSS.2018.XIV.049.
110. A. Kumar, A. Zhou, G. Tucker, S. Levine, "Conservative Q-learning for offline reinforcement learning" in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, 2020), pp. 1179–1191.
111. S. Fujimoto, S. Gu, "A minimalist approach to offline reinforcement learning" in *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, J. Wortman Vaughan, Eds. (Curran Associates, 2021), pp. 20132–20145.
112. S. Reddy, A. D. Dragan, S. Levine, "SQL: Imitation learning via reinforcement learning with sparse rewards" in *International Conference on Learning Representations (ICLR)* (ICLR, 2020), pp. 1–14.
113. D. Garg, S. Chakraborty, C. Cundy, J. Song, S. Ermon, "IQ-learn: Inverse soft-Q learning for imitation" in *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, J. Wortman Vaughan, Eds. (Curran Associates, 2021), pp. 4028–4039 (2021).
114. B. Keller, M. Draeos, K. Zhou, R. Qian, A. N. Kuo, G. Konidaris, K. Hauser, J. A. Izatt, Optical coherence tomography-guided robotic ophthalmic microsurgery via reinforcement learning from demonstration. *IEEE Trans. Robot.* **36**, 1207–1218 (2020).
115. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. arXiv:1707.06347 [cs.LG] (2017).
116. Z. Cui, J. Cartucho, S. Giannarou, F. Rodriguez y Baena, Caveats on the first-generation da Vinci Research Kit: Latent technical constraints and essential calibrations [Survey]. *IEEE Robot. Autom. Mag.* **32**, 2–17 (2023).
117. O. Özgüner, T. Shkurti, S. Huang, R. Hao, R. C. Jackson, W. S. Newman, M. C. Çavuşoğlu, Camera-robot calibration for the Da Vinci robotic surgery system. *IEEE Trans. Autom. Sci. Eng.* **17**, 2154–2161 (2020).
118. K. Pachtrachai, M. Allan, V. Pawar, S. Hailes, D. Stoyanov, "Hand-eye calibration for robotic assisted minimally invasive surgery without a calibration object" in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2016), pp. 2485–2491.
119. Z. Wang, Z. Liu, Q. Ma, A. Cheng, Y. H. Liu, S. Kim, A. Deguet, A. Reiter, P. Kazanzides, R. H. Taylor, Vision-based calibration of dual RCM-based robot arms in human-robot collaborative minimally invasive surgery. *IEEE Robot. Autom. Lett.* **3**, 672–679 (2018).
120. Y. Li, F. Richter, J. Lu, E. K. Funk, R. K. Orosco, J. Zhu, M. C. Yip, SuPer: A surgical perception framework for endoscopic tissue manipulation with surgical robotics. *IEEE Robot. Autom. Lett.* **5**, 2294–2301 (2020).
121. A. E. Abdelaal, J. Fang, T. N. Reinhart, J. A. Mejia, T. Z. Zhao, J. Bohg, A. M. Okamura, Force-aware autonomous robotic surgery. arXiv:2501.11742 [cs.LG] (2025).
122. Y. Wang, J. Wang, Y. Li, T. Yang, C. Ren, The deep reinforcement learning-based VR training system with haptic guidance for catheterization skill transfer. *IEEE Sens. J.* **22**, 2356–23366 (2022).
123. P. Puangmal, K. Althoefer, L. D. Seneviratne, D. Murphy, P. Dasgupta, State-of-the-art in force and tactile sensing for minimally invasive surgery. *IEEE Sens. J.* **8**, 371–381 (2008).
124. H. Ding, Y. Zhang, H. Shu, Y. Long, C. Gao, T. Lu, R. Liang, L. Seenivasan, Q. Dou, M. Unberath, SegSTRONG-C: Segmenting surgical tools robustly on non-adversarial generated corruptions—An EndoVis'24 challenge. arXiv:2407.11906 [cs.CV] (2024).
125. V. Penza, E. de Momi, N. Enayati, T. Chupin, J. Ortiz, L. S. Mattos, EnViSoRS: Enhanced vision system for robotic surgery. A user-defined safety volume tracking to minimize the risk of intraoperative bleeding. *Front. Robot. AI* **4**, 15 (2017).

126. Y. Ou, M. Tavakoli, CRESSim-MPM: A material point method library for surgical soft body simulation with cutting and suturing. *arXiv:2411.14622 [cs.RO]* (2024).
127. W. Chang, Y. Li, Z. Zhu, Y. Yang, "LSD3K: A benchmark for smoke removal from laparoscopic surgery images" in *2024 3rd International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology (AIoTC)* (IEEE, 2024), pp. 1–5.
128. D. Wang, J. Qi, B. Huang, E. Noble, D. Stoyanov, J. Gao, D. S. Elson, Polarization-based smoke removal method for surgical images. *Biomed. Opt. Express* **13**, 2364–2379 (2022).
129. T. Okamoto, T. Ohnishi, H. Kawahira, O. Dergachyava, P. Jannin, H. Haneishi, Real-time identification of blood regions for hemostasis support in laparoscopic surgery. *Signal Image Video Process.* **13**, 405–412 (2019).
130. W. Reiter, Co-occurrence balanced time series classification for the semi-supervised recognition of surgical smoke. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 2021–2027 (2021).
131. S. Reed, K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-Maron, M. Gimenez, Y. Sulsky, J. Kay, J. T. Springenberg, T. Eccles, J. Bruce, A. Razavi, A. Edwards, N. Heess, Y. Chen, R. Hadsell, O. Vinyals, M. Bordbar, N. de Freitas, A generalist agent. *arXiv:2205.06175 [cs.AI]* (2022).
132. A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, T. Jackson, S. Jesmonth, N. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, K.-H. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada, J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. Ryoo, G. Salazar, P. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, B. Zitkovich, RT-1: Robotics transformer for real-world control at scale. *arXiv:2212.06817 [cs.RO]* (2022).
133. A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dube, C. Finn, P. Florence, C. Fu, M. G. Arenas, K. Gopalakrishnan, K. Han, K. Hausman, A. Herzog, J. Hsu, B. Ichter, A. Irpan, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, L. Lee, Tsang-Wei Edward Lee, S. Levine, Y. Lu, H. Michalewski, I. Mordatch, K. Pertsch, K. Rao, K. Reymann, M. Ryoo, G. Salazar, P. Sanketi, P. Sermanet, J. Singh, A. Singh, R. Soricut, H. Tran, V. Vanhoucke, Q. Vuong, A. Wahid, S. Welker, P. Wohlhart, J. Wu, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, B. Zitkovich, RT-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv:2307.15818 [cs.RO]* (2023).
134. Open X-Embodiment Collaboration, Open X-Embodiment: Robotic learning datasets and RT-X models (2023); <https://robotics-transformer-x.github.io>.
135. Y. Hu, Q. Xie, V. Jain, J. Francis, J. Patrikar, N. Keetha, S. Kim, Y. Xie, T. Zhang, H. Fang, S. Zhao, S. Omidshafiei, D. Kim, A. Agha-Mohammadi, K. Sycara, M. Johnson-Roberson, D. Batra, X. Wang, S. Scherer, C. Wang, Z. Kira, F. Xia, Y. Bisk, Toward general-purpose robots via foundation models: A survey and meta-analysis. *arXiv:2312.08782 [cs.RO]* (2023).
136. K.-H. Oh, B. Leonardo, M. Alberto, V. Valentini, P. M. Di, T. Francesco, P. Gioia, A. Luciano, D. Alvaro, Z. Milos, C. Liaohai, P. C. Giulianotti, Comprehensive Robotic Cholecystectomy Dataset (CRCD): Integrating kinematics, pedal signals, and endoscopic videos. *arXiv:2312.01183 [cs.RO]* (2023).
137. Y. Gao, S. Vedula, C. Reiley, N. Ahmadi, B. Varadarajan, H. Lin, L. Tao, L. Zappella, B. Béjar, D. Yuh, C. Chen, R. Vidal, S. Khudanpur, G. Hager, JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling, paper presented at Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014 workshop: Modeling and Monitoring of Computer-Assisted Interventions, Boston, MA, 14 September 2014.
138. M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Raffailov, E. Foster, G. Lam, P. Sanketi, Q. Vuong, T. Kollar, B. Burchfiel, R. Tedrake, D. Sadigh, S. Levine, P. Liang, C. Finn, OpenVLA: An open-source vision-language-action model. *arXiv:2406.09246 [cs.RO]* (2024).
139. H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovitch, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, T. Scialom, Llama 2: Open foundation and fine-tuned chat models. *arXiv:2307.09288 [cs.CL]* (2023).
140. L. Zheng, W. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. Xing, H. Zhang, J. Gonzalez, I. Stoica, "Judging LLM-as-a-judge with MT-Bench and Chatbot Arena" in *Advances in Neural Information Processing Systems* **36**, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates, 2024), pp. 46595–46623.
141. D. Driess, F. Xia, M. S. M. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, W. Huang, Y. Chebotar, P. Sermanet, D. Duckworth, S. Levine, V. Vanhoucke, K. Hausman, M. Toussaint, K. Greff, A. Zeng, I. Mordatch, P. Florence, PaLM-E: An embodied multimodal language model. *arXiv:2303.03378 [cs.LG]* (2023).
142. Y. Li, Y. Huang, M. E. Ildiz, A. S. Rawat, S. Oymak, "Mechanics of next token prediction with self-attention" in *Proceedings of the International Conference on Artificial Intelligence and Statistics* (PMLR, 2024), pp. 685–693.
143. S. Zhang, L. Dong, X. Li, S. Zhang, X. Sun, S. Wang, J. Li, R. Hu, T. Zhang, F. Wu, G. Wang, Instruction tuning for large language models: A survey. *arXiv:2308.10792 [cs.CL]* (2023).
144. B. Peng, C. Li, P. He, M. Galley, J. Gao, Instruction tuning with GPT-4. *arXiv:2304.03277 [cs.CL]* (2023).
145. H. Liu, C. Li, Q. Wu, Y. J. Lee, "Visual instruction tuning" in *Advances in Neural Information Processing Systems* **36**, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates, 2024), pp. 34892–34916.
146. T. Buckley, J. A. Diao, A. Rodman, A. K. Manrai, Multimodal foundation models exploit text to make medical image predictions. *arXiv:2311.0559 [cs.CV]* (2023).
147. A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, D. S. W. Ting, Large language models in medicine. *Nat. Med.* **29**, 1930–1940 (2023).
148. Y. Chebotar, Q. Vuong, K. Hausman, F. Xia, Y. Lu, A. Irpan, A. Kumar, T. Yu, A. Herzog, K. Pertsch, K. Gopalakrishnan, J. Ibarz, O. Nachum, S. A. Sontakke, G. Salazar, H. T. Tran, J. Peralta, C. Tan, D. Manjunath, J. Singh, B. Zitkovich, T. Jackson, K. Rao, C. Finn, S. Levine, "Q-Transformer: Scalable offline reinforcement learning via autoregressive Q-functions" in *Proceedings of the 7th Conference on Robot Learning*, J. Tan, M. Toussaint, K. Darvish, Eds., vol. 229 of *Proceedings of Machine Learning Research* (PMLR, 2023), pp. 3909–3928.
149. M. Fontana, G. Zeni, S. Vantini, Conformal prediction: A unified review of theory and new challenges. *Bernoulli* **29**, 1–23 (2023).
150. J. Sun, Y. Jiang, J. Qiu, P. Nobel, M. Kochenderfer, M. Schwager, "Conformal prediction for uncertainty-aware planning with diffusion dynamics model" in *Advances in Neural Information Processing Systems* **36**, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine, Eds. (Curran Associates, 2024), pp. 80324–80337.
151. S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, I. S. Kohane, Adversarial attacks on medical machine learning. *Science* **363**, 1287–1289 (2019).
152. X. Ma, Y. Niu, L. Gu, Y. Wang, Y. Zhao, J. Bailey, F. Lu, Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognit.* **110**, 107332 (2021).
153. K. R. Ludvigsen, S. Nagaraja, Dissecting liabilities in adversarial surgical robot failures: A national (Danish) and EU law perspective. *Comput. Law Secur. Rev.* **44**, 105656 (2022).
154. N. Ghaffari Laleh, D. Truhn, G. P. Veldhuizen, T. Han, M. van Treeck, R. D. Buelow, R. Langer, B. Dislich, P. Boor, V. Schulz, J. N. Kather, Adversarial attacks and adversarial robustness in computational pathology. *Nat. Commun.* **13**, 5711 (2022).
155. M. Sadi, B. M. S. B. Talukder, K. Mishty, M. T. Rahman, Attacking deep learning AI hardware with universal adversarial perturbation. *Information* **14**, 516 (2023).
156. V. Muralidharan, B. A. Adewale, C. J. Huang, M. T. Nta, P. O. Ademiju, P. Pathmarajah, M. K. Hang, O. Adesanya, R. O. Abdullateef, A. O. Babatunde, A. Ajibade, S. Onyekia, Z. R. Cai, R. Daneshjou, T. Olatunji, A scoping review of reporting gaps in FDA-approved AI medical devices. *NPI Digit. Med.* **7**, 273 (2024).
157. B. Yacoub, A. Varga-Szemes, U. J. Schoepf, I. M. Kabakus, D. Baruah, J. R. Burt, G. J. Aquino, A. K. Sullivan, J. O'Doherty, P. Hoelzer, J. Sperl, T. Emrich, Impact of artificial intelligence assistance on chest CT interpretation times: A prospective randomized study. *Am. J. Roentgenol.* **219**, 743–751 (2022).
158. A. A. van der Heijden, M. D. Abramoff, F. Verbraak, M. V. van Hecke, A. Liem, G. Nijpels, Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol.* **96**, 63–68 (2018).
159. FDA, "Artificial intelligence-enabled device software functions: Lifecycle management and marketing submission recommendations: Draft guidance for industry and Food and Drug Administration staff" (FDA-2024-D-4488, 2025); <https://fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing>.
160. H. Al Hajj, M. Lamard, P.-H. Conze, B. Cochener, G. Quelled, Monitoring tool usage in surgery videos using boosted convolutional and recurrent neural networks. *Med. Image Anal.* **47**, 203–218 (2018).
161. K. Schoeffmann, M. Taschwer, S. Sarny, B. Münzer, M. Jürgen Primus, D. Putzgruber, "Cataract-101: Video dataset of 101 cataract surgeries" in *Proceedings of the 9th ACM Multimedia Systems Conference* (ACM, 2018), pp. 421–425.
162. D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele, P. Jannin, Detecting surgical tools by modelling local appearance and global shape. *IEEE Trans. Med. Imaging* **34**, 2603–2617 (2015).
163. A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, N. Padoy, EndoNet: A deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans. Med. Imaging* **36**, 86–97 (2017).
164. W.-Y. Hong, C.-L. Kao, Y.-H. Kuo, J.-R. Wang, W.-L. Chang and C.-S. Shih, CholecSeg8k: A semantic segmentation dataset for laparoscopic cholecystectomy based on Cholec80. *arXiv:2012.12453 [cs.CV]* (2020).
165. C. I. Nwoye, T. Yu, C. Gonzalez, B. Seeliger, P. Mascagni, D. Mutter, J. Marescaux, N. Padoy, Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Med. Image Anal.* **78**, 102433 (2022).
166. N. Madapana, M. Rahman, N. Sanchez-Tamayo, M. Balakuntala, G. Gonzalez, J. Bindu, V. Venkatesh, X. Zhang, J. Noguera, T. Low, R. Voyles, Y. Xue, J. Wachs, "DESK: A robotic activity dataset for dexterous surgical skills transfer to medical robots" in *2019 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS) (IEEE, 2019), pp. 6928–6934.

167. A. Huaulmé, K. Harada, Q.-M. Nguyen, B. Park, S. Hong, M.-K. Choi, M. Peven, Y. Li, Y. Long, Q. Dou, S. Kumar, S. Lalithkumar, R. Hongliang, H. Matsuzaki, Y. Ishikawa, Y. Harai, S. Kondo, M. Mitsuishi, P. Jannin, PEg TRAnsfer Workflow recognition challenge report: Does multi-modal data improve recognition? arXiv:2202.05821 [cs.LG] (2022).
168. E. D. Goodman, K. K. Patel, Y. Zhang, W. Locke, C. J. Kennedy, R. Mehrotra, S. Ren, M. Y. Guan, M. Downing, H. W. Chen, J. Z. Clark, G. A. Brat, S. Yeung, A real-time spatiotemporal AI model analyzes skill in open surgical videos. arXiv:2112.07219 [cs.CV] (2021).
169. Open X-Embodiment Collaboration, “Open X-embodiment: Robotic learning datasets and RT-X models,” presented at the workshop Towards Generalist Robots: Learning Paradigms for Scalable Skill Acquisition at the 2023 Conference on Robot Learning, Atlanta, GA, 6 to 9 November 2023.

Acknowledgments

Funding: This material is based on work supported by the National Science Foundation under grants DGE 2139757 and NSF/FRR 2144348; NIH grant R56EB033807; and ARPA-H grants 75N91023C00048, AY1AX000023, and D24AC00415. **Author contributions:** S.S., J.O., J.W.K., and A.K. conceptualized this paper; S.S., J.O., and J.W.K. wrote the paper; S.S., J.O., J.W.K., and A.K. edited the paper; A.K. provided funding. **Competing interests:** A.K. is an inventor on a patent application related to autonomous surgery: “Automated Surgical and Interventional Procedures,” US Patent application US-2014-0005684-A1.

Submitted 8 September 2024

Accepted 27 June 2025

Published 23 July 2025

10.1126/scirobotics.adt0187

Will your next surgeon be a robot? Autonomy and AI in robotic surgery

Samuel Schmidgall, Justin D. Opfermann, Ji Woong Kim, and Axel Krieger

Sci. Robot. **10** (104), eadt0187. DOI: 10.1126/scirobotics.adt0187

View the article online

<https://www.science.org/doi/10.1126/scirobotics.adt0187>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works