



The future of robotic artificial intelligence in acute care surgery: an early proving ground and a final frontier for autonomous surgical robotics

Joel Grunhut¹, Ruhika Iyer², Jesse Haworth³, Gabriel Brat⁴, Axel Krieger³, Jeffrey K. Jopling⁵

Keywords:

Acute care surgery, surgical robotics, autonomous surgery, artificial intelligence, autonomy

Citation: Grunhut J, Iyer R, Haworth J, Brat G, Krieger A, Jopling JK. The future of robotic artificial intelligence in acute care surgery: an early proving ground and a final frontier for autonomous surgical robotics. *Art Int Surg*. 2026;6:541-55.

<https://dx.doi.org/10.20517/ais.2026.74>

Received: 19 July 2026

First Decision: 26 Aug 2026

Revised: 30 Sep 2026

Accepted: 10 Oct 2026

Published: 23 Oct 2026

Academic Editors:

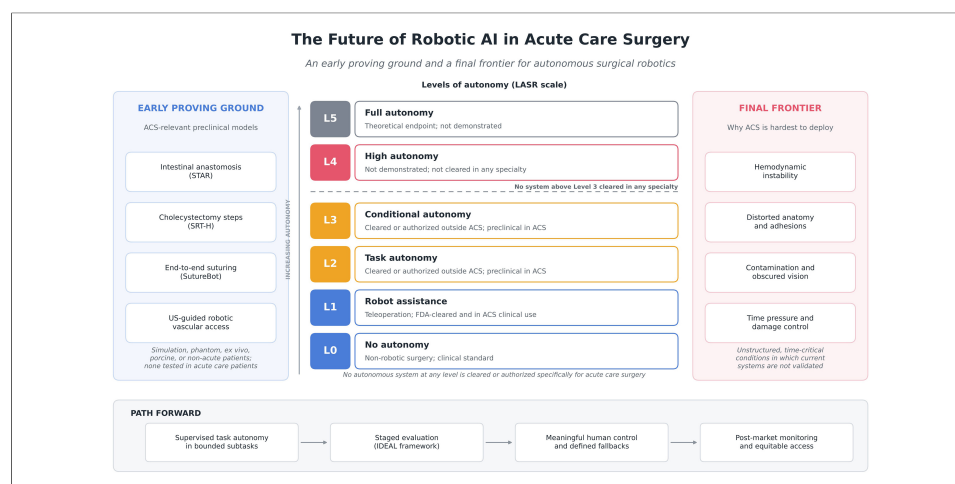
Andrew Gumbs, Belinda De Simone

Copy Editor:

Tong Wang

Production Editor:

Tong Wang



Abstract

Acute care surgery (ACS), spanning emergency general surgery, trauma surgery, and surgical critical care, demands high-stakes decisions in settings characterized by hemodynamic instability, distorted anatomy, contamination, and time pressure. Artificial intelligence (AI) is proposed as a means to reduce cognitive burden, standardize technically sensitive tasks, and extend surgical capability in resource-constrained environments. This review examines the autonomy continuum in surgical AI, with ACS as the primary lens. We argue that ACS occupies a paradoxical position: it has produced the earliest preclinical demonstrations of autonomous surgical robotics, yet the very conditions that motivate autonomy in this specialty make it among the most challenging environments in which to deploy these systems safely in human patients. For ACS, the central question is less whether AI can perform surgical tasks than how autonomy should be sequenced and bounded for emergency surgery. We synthesize the literature across four domains: (1) how autonomy is defined and operationalized in surgical AI and robotics; (2) the current and emerging spectrum of automation in operative environments; (3) the

¹Department of Surgery, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

²Department of Biology, New Jersey Institute of Technology, Newark, NJ 07102, USA.

³Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

⁴Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA.

⁵Department of Surgery, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA.

Correspondence to: Dr. Joel Grunhut, Department of Surgery, University of Maryland School of Medicine, Baltimore, MD 21201, USA. E-mail: jgrunhut@som.umaryland.edu

technical feasibility of automation for ACS-relevant tasks; and (4) the ethical boundaries of autonomy in high-risk, time-critical cases. We argue that the safest course is incremental: deploying supervised task autonomy where safety properties are demonstrable and reserving higher autonomy for rigorously evaluated, tightly bounded domains. Realizing this course will require evaluation strategies that align autonomy level to clinical risk, preserve meaningful human oversight, support post-market monitoring, and address liability, patient trust, and equitable access.

INTRODUCTION

Surgeons in acute care settings routinely make consequential decisions with incomplete information, under time pressure, and in the face of rapidly evolving physiology. Emergency general surgery and trauma care are defined by this unpredictability: the patient with a perforated viscus who arrives hypotensive, the trauma laparotomy where anatomy is obscured by hemorrhage, the critical care escalation where competing interventions must be prioritized within minutes. Classic rule-based decision-support systems are limited chiefly by rigidity: they encode simplified, static assumptions and degrade when inputs fall outside predefined logic. Machine-learning and artificial intelligence (AI) systems overcome some of this rigidity but introduce a distinct set of risks: dataset bias, performance drift under distribution shift, and failure modes that are less transparent and harder to anticipate^[1,2].

A central question for acute care surgery (ACS) is therefore not simply whether AI “works”, but where it should sit on the autonomy spectrum: tools that advise, tools that act with supervision, and tools that execute independently. That spectrum is now visible in both clinical software (e.g., risk calculators that inform triage) and embodied systems (e.g., robotic platforms that perform physical tasks), and it is increasingly shaped by ethical imperatives and evaluation frameworks as much as by technical feasibility. Throughout, we distinguish AI-based decision support (systems that analyze data and advise) from embodied robotic autonomy (systems that physically execute operative tasks); the two raise different safety, evaluation, and regulatory questions and should not be conflated^[3,4].

This creates a paradox. ACS-relevant procedures have served as experimental models for some of the most ambitious demonstrations of surgical autonomy to date: autonomous laparoscopic intestinal anastomosis, hierarchical frameworks for cholecystectomy, and end-to-end suturing benchmarks have all emerged from ACS-adjacent research, performed under controlled, preclinical conditions (simulation, *ex vivo*, and animal models); none has been tested in acute care patients^[5-7]. Yet the operative conditions that define emergency surgery (hemodynamic instability, contaminated fields, adhesions, tissue friability, and the need for rapid escalation to damage control) represent the hardest unsolved problems in surgical robotics. ACS may therefore occupy a unique position: among the earliest specialties to inspire laboratory demonstrations of autonomous capability through ACS-relevant procedures, yet potentially among the last in which such systems can be deployed safely in human patients. We advance this first-and-last framing as a conceptual hypothesis. It makes no historical or predictive claim. Understanding this tension, and charting a responsible path through it, is the central aim of this review.

METHODS

We conducted a targeted narrative review of operative ACS-related AI, focusing on partial automation and autonomous execution, technical feasibility, and ethical constraints. We searched PubMed and Google Scholar and hand-searched reference lists of key articles, using combinations of the concepts “autonomy/autonomous”, “surgical robotics”, “acute care/emergency/trauma surgery”, “machine learning/AI”, “validation”, “ethics”, and “regulation”, prioritizing publications from 2018 to 2026. The final search was conducted on 5 March 2026; one subsequently published, directly relevant source (cited at first

use in Part 2) was added afterward as a targeted addition identified by the author group. Peer-reviewed clinical and robotics literature was prioritized; preprints, Food and Drug Administration (FDA) guidance, and other regulatory documents were included only where peer-reviewed data were not yet available, and are identified as such in the reference list. Because direct evidence in acute care patients is limited, we deliberately included ACS-adjacent and preclinical work (e.g., autonomous ultrasound-guided vascular access) that points to near-term directions for the specialty. Where possible, we indicate the predominant evidence type (simulation, phantom, *ex vivo*, animal, or human) for each cited demonstration, applied pragmatically by the author group without a formal, protocolized adjudication procedure. Autonomy levels referenced throughout follow the frameworks cited in Part 1. The task-autonomy readiness matrix [Table 1] and the translational-gap analysis [Table 2] were developed by the authors: candidate ACS subtasks and gaps were enumerated, mapped to the identified evidence, and assigned target autonomy levels, failure modes, and fallback behaviors through iterative discussion among the multidisciplinary author group. They represent a proposed expert framework; they have not been formally validated or subjected to consensus, and constitute the principal original contribution of this review.

PART 1. DEFINING AUTONOMY IN ACS AI

A level-based framework

Discussions of surgical AI are often clouded by the failure to distinguish between levels of autonomy. Level-based frameworks for human-automation interaction predate surgical robotics: Parasuraman and colleagues' model of automation, developed for aviation and process control, established the underlying logic that higher levels progressively shift information acquisition, analysis, decision selection, and action implementation from the human to the machine^[8]. To orient the discussion for surgery specifically, we adopt a level-based view of surgical autonomy adapted from recent regulatory and robotics literature [Figure 1]^[4,9-11].

What exists today versus what has been demonstrated

Research demonstrations must be distinguished from clinical reality. The FDA's AI-Enabled Medical Devices list, which as of 2025 includes over 1,000 authorized devices, illustrates this gap^[12]. As of this writing, no systems higher than Level 3 have been cleared or authorized for clinical use in any surgical specialty, and no autonomous system of any level has been cleared or authorized specifically for ACS^[4]. The vast majority of FDA-cleared surgical robots remain at the lowest autonomy levels (predominantly Level 1), providing teleoperation with motion scaling, tremor filtering, or virtual fixtures^[4]. The highest currently cleared autonomy is Level 3, comprising three systems: TSolution One (Think Surgical) for autonomous bone milling in knee and hip arthroplasty, iSR'obot Mona Lisa 2.0 (Biobot) for transperineal prostate biopsy, and the ARTAS iX System (Venus Concept) for robotic follicular unit harvesting; AquaBeam (Procept BioRobotics) for prostate resection is classified at Level 2^[4]. Critically, all operate on rigid or predictable anatomy with well-defined geometric targets, conditions fundamentally different from the inflamed bowel, adhesions, and hemodynamic instability of emergency surgery.

Scientific rigor and the implementation gap

Even where model performance appears strong on development datasets, translation to clinical practice remains uneven. Loftus and colleagues' 2023 review emphasizes that many surgical decision-support models exhibit deficiencies in scientific rigor and readiness for clinical use, including limited external validation and inconsistent reporting of clinically meaningful performance characteristics^[13,14]. Parallel themes appear in broader surgical AI guidance, which stresses that surgeons must understand model limitations, dataset provenance, and operational constraints before adopting these tools^[15,16].

ACS introduces specific pitfalls that compound these concerns. Time-dependent confounding, arising from the rapid physiologic changes characteristic of ACS patients, can invalidate predictions made moments earlier. Inequitable model performance becomes a patient-safety issue when social and structural

Table 1. ACS task autonomy readiness matrix

Task	Target autonomy	Current best evidence	Uncertainty sources	Key metrics	Principal failure modes	Proposed fallbacks
Vascular access (US-guided)	L2 (L3 research target)	Phantom, porcine (shock models); early human feasibility (central venous access, not in acute care patients) ^[19-21]	Vessel collapse (hypotension), aberrant anatomy, image noise, patient movement, supervisor divided attention during resuscitation	Needle-force limit, takeover latency, complication rate vs. expert, OOD detection with pause/retract/handover	Vessel miss, needle misdirection, hematoma	Pause, retract needle, alert surgeon, handover
Camera/View control	L2	Phantom ^[58]	Blood/smoke occlusion, instrument collision, target ambiguity, rapid scene changes, automation complacency during sustained autonomous view	Contact-force limit, takeover latency, FOV/visibility vs. expert, OOD detection with return-to-view/handover	Loss of target, collision, delayed response to occlusion	Pause, return to last-known-good view, handover
Suction	L1-L2	Simulation ^[48]	Fluid volume variability, tissue proximity, unpredictable bleeding sources, delayed recognition of field obscuration under workload	Tip-force limit, takeover latency, clearance efficacy vs. expert, OOD detection with retract/takeover	Tissue contact injury, inadequate clearing	Retract, pause, manual takeover
Exposure/Retraction	L2-L3	Simulation, ex vivo, limited in vivo ^[37] [preprint]	Tissue stiffness variability (edema, fibrosis, inflammation), force estimation error, automation complacency regarding force limits	Tissue-tension limit, takeover latency, slip rate vs. expert, OOD detection with release-tension/pause/alert surgeon	Tissue tear, excessive force, instrument clash	Release tension, pause, alert surgeon
Suturing/Anastomosis	L3	Simulation, phantom, ex vivo, porcine ^[5,7,25]	Tissue friability, needle-tissue interaction variability, anatomic distortion, divided attention during mid-stitch autonomous execution	Force limit, takeover latency (mid-stitch), leak/knot security vs. expert, OOD detection with pause/handover	Needle drop, tissue tear, misalignment	Pause mid-stitch, request confirmation, handover
Autonomous suction/field clearing	L2	Simulation ^[48]	Bleeding source localization uncertainty, energy dosing variability, tissue charring effects, delayed recognition of escalating blood loss under divided attention	Thermal limit, takeover latency, hemostasis completeness vs. expert, OOD detection with pause/handover	Thermal spread, incomplete hemostasis, wrong target	Pause, alert surgeon, handover

This table maps current evidence, uncertainty sources, key performance metrics, principal failure modes, and proposed fallback behaviors to six representative ACS subtasks. Target autonomy levels reflect preclinical feasibility and represent the authors' proposed assessment based on the cited preclinical literature and expert judgment, without formal consensus. No systems higher than Level 3 have been cleared or authorized for clinical use, and no autonomous system of any level has been cleared or authorized specifically for ACS. Key Metrics for each task are organized around four safety-property categories: a task-specific force or dose limit, surgeon takeover latency, an outcome measure benchmarked for noninferiority against expert performance, and out-of-distribution detection coupled to successful fallback execution. These are author-proposed target categories aligned with the operational definition of demonstrable safety properties in the roadmap section above; validated numeric thresholds have not yet been established in the literature and must be derived through staged (IDEAL) evaluation^[18,43,44]. FOV: Field of view; L: Level; OOD: out-of-distribution; US: ultrasound; ACS: acute care surgery; IDEAL: Idea, Development, Exploration, Assessment, and Long-term monitoring.

Table 2. Key translational gaps for ACS autonomy

Domain	Gap	Path Forward
Data	Limited annotated ACS datasets; underrepresentation of pathological anatomy (hostile abdomen, adhesions, contamination)	Federated learning across trauma centers; synthetic data augmentation with domain randomization; prospective ACS-specific registries
Validation	Current evidence limited to simulation, phantoms, and porcine models; no human data for Level 3+ tasks ^[43]	Staged progression: simulation stress tests → calibrated phantoms → animal shock models → supervised first-in-human trials
Uncertainty quantification	Most models lack calibrated confidence estimates; no standard for out-of-distribution detection in surgical AI ^[14]	Require uncertainty quantification (ensemble, conformal prediction) as standard for regulatory submission; validate under ACS stress conditions
Fallback protocols	No validated fallback protocols for ACS autonomy failures ^[44]	Define and stress-test pause/retract/handover behaviors for each task [see Table 1] before clinical deployment
Regulation	No FDA pathway specific to ACS autonomy; Class II 510(k) may be insufficient for Level 3+ ^[44,53]	Engage FDA early; consider PMA pathway for high-risk ACS applications; develop ACS-specific predetermined change control plans
Consent	Emergency context complicates standard informed consent for some first-in-human trials	Assess applicability of the 21 CFR 50.24 exception on a case-by-case basis; where it applies, initiate consultation early and establish independent data monitoring

This table identifies critical translational gaps separating preclinical demonstrations from clinical deployment and proposes paths forward across six domains. ACS: Acute care surgery; AI: artificial intelligence; FDA: Food and Drug Administration; OOD: out-of-distribution; PMA: premarket approval.

determinants influence both how patients present and how they are treated. If AI systems are deployed as opaque authorities instead of transparent tools, bias and over-reliance become direct threats to care quality^[1,17].

PART 2. PARTIAL AUTOMATION: BOUNDED ACTIONS UNDER OVERSIGHT

Partial automation describes systems that complete specific tasks or subtasks while a clinician retains supervisory control and the ability to intervene^[18]. At Autonomy Levels 2 and 3, task autonomy and supervised autonomy often target repetitive, time-consuming, or ergonomically difficult tasks. These are the most reachable next step for ACS because they target high-frequency procedural bottlenecks without demanding full procedural autonomy.

Automation adjacent to ACS: vascular access and imaging

A relevant ACS example of Level 2 autonomy is AI-enabled, ultrasound-guided robotic vascular access, a task with clear time sensitivity during resuscitation and frequent dependence on operator skill. Brattain and colleagues described a handheld system combining ultrasound guidance and robotic actuation to facilitate femoral access. Here partial automation amplifies and standardizes operator skill; it does not replace the operator^[19]. Automated vascular access currently operates at Level 2, executing a constrained sequence under continuous operator control; more advanced machine-learning models are under investigation to expand adaptive capability toward supervised conditional autonomy^[19-21].

Conditional autonomy in minimally invasive tasks

Level 3 partial automation is increasingly appearing in the robotic surgery literature as conditional autonomy, where the system executes defined operative steps while humans supervise and manage exceptions^[22]. A representative example is conditional autonomy for robotic appendectomy, which explicitly positions autonomy as incremental and task-scoped; independent surgery is not the near-term goal^[23]. Similarly, semi-autonomous gallbladder retraction demonstrates a realistic near-term form of automation:

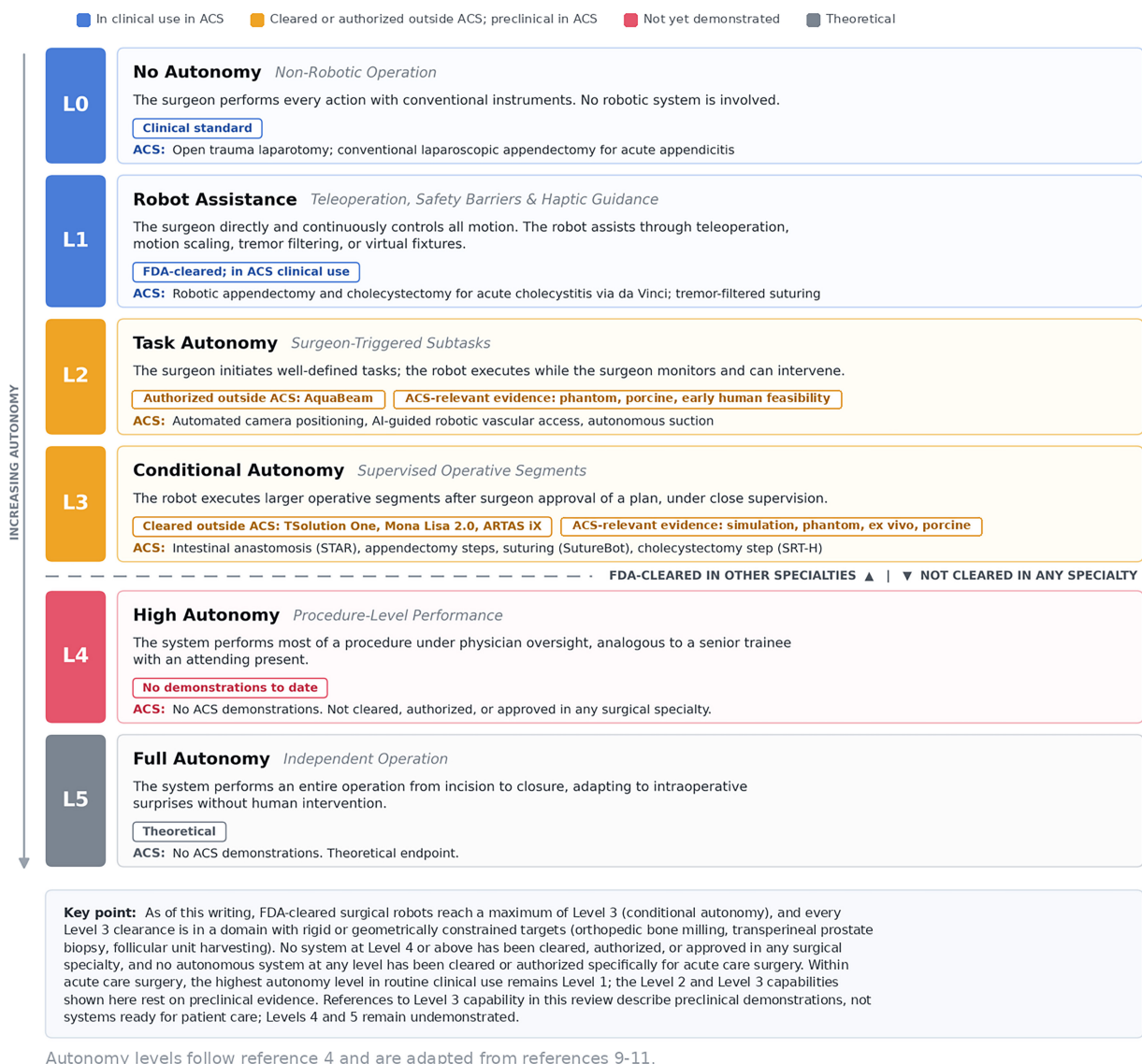


Figure 1. Levels of autonomy in surgical robotics, from Level 0 (no autonomy; equivalent to non-robotic, manual surgery) through Level 5 (full autonomy). Teleoperated robotic surgery, in which the surgeon continuously controls all motion and the system provides passive support such as motion scaling or tremor filtration, is classified as Level 1 (robot assistance). At each ascending level, the system assumes greater responsibility for perception, planning, and action while human control correspondingly decreases. Levels follow the LASR scale of Lee et al.^[4] and are adapted from the surgical-autonomy frameworks of Haidegger, Yang et al., and Opfermann et al.^[9-11] Published frameworks differ on where teleoperation is classified: some, such as that of Yang et al.^[10], place teleoperated robots at Level 0, whereas the LASR scale reserves Level 0 for non-robotic surgery and assigns teleoperation to Level 1. FDA device classifications shown in the figure, including the Level 3 systems TSolution One, iSR'obot Mona Lisa 2.0, and ARTAS iX, are those reported by Lee et al.^[4] Evidence labeled ACS-relevant was obtained in preclinical models or in patients outside the acute care setting, not in acute care patients. No system above Level 3 has been cleared or authorized for clinical use in any surgical specialty. ACS: Acute care surgery; FDA: Food and Drug Administration; LASR: Levels of Autonomy in Surgical Robotics.

exposure and tissue handling performed reliably in models. This frees surgeons to focus on higher-order judgment while preserving human oversight for critical safety decisions^[24].

PART 3. AUTONOMOUS EXECUTION: EXPERIMENTAL BUT RAPIDLY EVOLVING

ACS at the frontier of autonomous surgery demonstrations

A meaningful share of the autonomous-execution demonstrations cited in this section rest on preprints not yet peer-reviewed; we flag the most consequential of these inline. Some of the most ambitious autonomous surgery demonstrations to date have emerged from ACS-relevant procedures. The Smart Tissue

Autonomous Robot (STAR) framework demonstrated Level 3 autonomous laparoscopic intestinal anastomosis in phantom, *ex vivo*, and animal (porcine) models, combining fluorescent imaging, 3D visual tracking, and force sensing to achieve dexterity comparable to expert surgeons; these are ACS-relevant preclinical demonstrations, not ACS clinical evidence, and no acute care patients were studied^[5,25]. This work suggests that, with further validation in contaminated and inflamed bowel typical of acute abdomen cases, autonomous suturing could eventually support ACS teams managing perforated viscus, strangulated hernias, or traumatic small bowel injuries^[26]. The Hierarchical Surgical Robot Transformer (SRT-H) extended this line of work, combining language-conditioned planning with low-level control to perform a full surgical step in *ex vivo* cholecystectomy models^[6]. Unlike prior systems that replayed fixed trajectories, SRT-H could adapt its behavior in response to intraoperative variation; whether this adaptability extends to ACS-type anatomical variation and hemodynamic instability, rather than the elective *ex vivo* model in which it was demonstrated, remains untested. A Level 3 autonomous suturing benchmark on the da Vinci Research Kit, spanning needle pickup, tissue insertion, and knot tying, has further demonstrated task-level capability^[7,27,28].

Technical feasibility: perception, learning, and integration

Developing perception and feedback for ACS

Autonomous execution in ACS depends on two capabilities: visual perception and sensory feedback^[29,30]. ACS procedures are most commonly performed either through open midline incisions with no natural video capture or with standard laparoscopic instruments that were not originally designed for force feedback. Robotic appendectomy and cholecystectomy will likely need to serve as the initial cases where video capture and haptic sensing can be captured and integrated. These will then serve as the foundation for training models that can handle the tissue variability typical of emergency cases: edematous bowel, inflamed mesentery, and anatomically distorted fields^[17].

Simulation-to-real transfer

Because collecting large-scale annotated datasets from live ACS cases is slow, expensive, and ethically constrained, simulation-based training with transfer to real operative environments (sim-to-real) is an important strategy. Emerging methods can now reconstruct three-dimensional tissue geometry from endoscopic images^[31], build physics models of deformable tissue from real surgical video^[32], simulate tissue cutting with realistic mechanical behavior^[33], and train autonomous agents entirely within simulated operative environments before transferring learned behaviors to real robots^[34,35].

However, sim-to-real transfer poses particular challenges for ACS. The same pathological conditions that create clinical urgency for automation (contamination, inflammation, tissue friability, hemodynamic instability) are precisely those that current simulators struggle to model faithfully. Soft-tissue deformation and stiffness vary unpredictably with pathology. Tissue planes may be obliterated in reoperative or contaminated abdomens. Models trained on “clean” simulations risk brittle performance when deployed in the reality of emergency surgery.

Future sim-to-real pipelines for ACS must incorporate domain randomization over tissue properties and pathological conditions, physics-accurate deformation models, and staged validation progressing from simulation through phantom, animal, and first-in-human testing.

Autonomous exposure, retraction, and force estimation

Exposure and retraction are rate-limiting steps in many ACS procedures. Autonomous systems that maintain optimal tissue tension using visual feedback, kinematic data, and reinforcement learning could offload repetitive manipulation while preserving surgeon oversight for high-risk maneuvers^[36]. Systems like

MEDiC, which autonomously maintain optimal surgical exposure, are potentially applicable, pending validation, to ACS teams working in narrow or dynamically changing fields of view^[37].

Vision-based force estimation methods have reported errors below 10% of applied force, without embedded sensors, in benchtop and *ex vivo* evaluations. This could help address the space and sterilization constraints of emergency surgery instruments. These figures reflect controlled experimental conditions and defined tissue models; none reflect clinically validated performance, and their relevance to ACS tissues remains unproven^[38]. These capabilities, combined with miniature force-sensing tools and automatic traction controllers, represent a pathway toward semi-autonomous assistance in exposure tasks during urgent laparotomy and laparoscopy^[39-41]. One caveat applies: these estimators must be validated against ACS-specific tissue properties (inflamed bowel, friable mesentery, hemorrhage-distorted anatomy) before safety claims can be made for ACS deployment.

Uncertainty-aware autonomy: a prerequisite for ACS deployment

For ACS, technical feasibility cannot be defined solely by whether a model can perform a task. It must also encompass whether the model can quantify its own confidence and act safely when that confidence is low. Modern end-to-end systems in autonomous driving provide a useful analogy. A perception stack identifies objects and free space, a planning stack outputs an intended trajectory, and a low-level controller executes while continuously checking for deviations. A similar multi-layered architecture is needed for ACS autonomy. Before executing a maneuver, the system should communicate what it is about to do, an associated confidence estimate, and an indication of whether current visual and force patterns fall within or outside its training distribution. When the system detects an out-of-distribution scenario (dense adhesions, unexpected bleeding, distorted anatomy), it should automatically alert the supervising surgeon and trigger a predefined fallback: pause, retract to a safe pose (a predefined low-risk instrument configuration), pack the wound, or hand back full control. This framing echoes two of the concrete safety research problems Amodei and colleagues identified for autonomous systems generally: robustness to distributional shift, and safe exploration under uncertainty [preprint]^[42].

A pragmatic roadmap for ACS autonomy

As autonomy increases, evaluation must extend beyond accuracy metrics to encompass human factors, failure-mode characterization, training and credentialing requirements, and real-world monitoring. The IDEAL (Idea, Development, Exploration, Assessment, and Long-term monitoring) Robotics framework offers a staged approach for development, comparative evaluation, and long-term surveillance that is directly applicable as ACS systems progress from prototypes to trials and post-market use^[18].

The task-specific readiness assessment in Table 1 complements this framework by mapping current evidence, uncertainty sources, failure modes, and proposed fallback behaviors to individual ACS subtasks. It identifies where bounded autonomy is most tractable and where the gap to safe deployment remains widest. Moving from qualitative readiness assessment to clinical trials will require defining quantitative performance targets for each task: acceptable complication rates relative to expert surgeon benchmarks, maximum allowable time for the surgeon to resume full control, and noninferiority margins that account for ACS patient heterogeneity. Together, the IDEAL framework and the task-readiness matrix offer a structured approach to sequencing investment and regulatory engagement across the autonomy spectrum.

For ACS, a demonstrable safety property is a pre-registered, quantitative, task-specific acceptance criterion, a higher bar than a general assurance that a system performed well in testing. No such criterion has yet been validated in a human study of a Level 3 system: a 2023 scoping review found all 129 intraoperative AI applications then published were still at IDEAL stage 0, with no human clinical investigations^[43], and a 2026

Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) white paper names the absence of clinically meaningful performance metrics as one of the field's central unmet regulatory challenges^[44]. What exists instead are endpoints from adjacent, better-established work. At the preclinical tier, the STAR intestinal-anastomosis platform reports the most concrete quality endpoints likely to carry into clinical evaluation: suture spacing and bite depth, anastomotic leak pressure, needle-placement corrections, completion time, and lumen patency, measured in phantom and *in vivo* porcine models with no acute care or human patients studied^[5,25]. At the clinical tier, the first-in-human Versius trial, though a teleoperated rather than autonomous platform, shows what an early conditional-autonomy study's safety endpoints would likely look like: unplanned conversion, return to the operating room, and 30- and 90-day readmission, reported in a Stage 2a interim analysis built toward IDEAL-D Stage 2b^[45]. Building on this evidence base, we propose candidate targets here as illustrative starting points, each still pending task-specific validation: a force- or tissue-tension limit set below the *ex vivo* damage threshold for the specific tissue and task; a complication or leak rate noninferior to an expert-surgeon benchmark, with a margin pre-specified along the lines the CORE-MD consortium recommends for high-risk device evaluation^[46]; and, for out-of-distribution scenarios, a validated detection rate paired with a required rate of successful fallback execution for each task in Table 1. Takeover latency deserves its own, narrower target: the time from an out-of-distribution trigger to restored surgeon control is a different clock than the communication delay studied in teleoperation, where performance measurably degrades above roughly 300 ms and becomes difficult above 600 to 700 ms^[47]. Deriving any of these values is itself a research task, properly sequenced through the IDEAL framework's staged evaluation.

Consistency and skill equalization

The goal of robotic autonomy in ACS is not to imitate surgical performance but to amplify expert-level technical consistency. In settings where surgical expertise varies, whether due to training differences, fatigue, case volume, or geographic isolation, structured autonomy could reduce variability in technically sensitive subtasks by embedding learned patterns from experienced surgeons into systems that execute bounded tasks with precision under supervision.

The earliest autonomous capabilities to reach ACS practice will likely function as skill equalizers: a camera-holding arm that maintains exposure as reliably at 3 a.m. in a rural trauma center as it does during a daytime case at an academic medical center, or a suturing assistant that achieves consistent bite spacing regardless of operator fatigue. They narrow the gap between ideal and real-world operative performance without replacing surgical judgment. Full autonomous capability, the kind that could independently navigate a hostile abdomen, remains the distant end of the spectrum and is, at present, speculative^[48,49]. But bounded autonomy that raises the floor of technical performance across diverse care settings represents a meaningful and achievable near-term contribution.

PART 4. ACCOUNTABILITY, ETHICS, AND THE PATIENT-SURGEON RELATIONSHIP

From recommendation to action: a shift in ethical risk

When AI systems transition from recommending to acting, ethical risk increases, because errors become immediate physical harms. They gain the ability to harm the patient on their own, whereas previously they only misinformed the surgeon. In ACS, physiologic instability and time pressure intensify this shift. The ethical bar therefore rises from "keep a clinician in the loop" to specifying meaningful human control in operational terms: what the surgeon must be able to perceive (system mode, intent, uncertainty), how quickly takeover can occur, and what the system does when uncertainty or patient status changes (pause, safe-state, request confirmation). Operationally, this entails specifying a maximum acceptable takeover time, the minimum information continuously visible to the supervising surgeon, and explicit authority to override or terminate system action. It also requires assigning responsibility for case selection and contraindications,

documenting the autonomy mode used in each case, and predefining accountability when failures arise from software updates, sensor errors, or inappropriate deployment. Autonomy level becomes a patient-safety variable rather than a marketing descriptor, and low-to-moderate autonomy is more defensible in ACS unless reliable interruption and fallback have been proven^[2,17].

The responsibility gap

Systems that act reshape accountability in ways that matter uniquely in ACS. Supervision is often intermittent during resuscitation. Competing priorities demand divided attention and the patient may deteriorate faster than a human can diagnose the automation's failure mode. This creates a predictable responsibility gap unless institutions define accountability structures for case selection and contraindications, supervisory competence and credentialing, and escalation rules. A useful starting apportionment distinguishes three loci of responsibility for intraoperative harm: manufacturer or algorithm-developer liability for device defects, inadequate validation, or unsafe autonomous behavior within the system's cleared envelope; institutional liability for credentialing decisions, case selection, and deployment outside validated indications; and supervising-surgeon responsibility for intraoperative oversight and timely takeover, consistent with an emerging framework for risk and liability in AI-enabled surgical systems^[50]. From a litigation-aware standpoint, ethical deployment requires a device-like evidentiary posture: version control, explicit documentation of the autonomy mode in use, and high-fidelity audit trails sufficient for root-cause analysis and transparent disclosure after harm, the foundation on which fair apportionment across these three loci actually depends.

Trust and the emergency patient-surgeon relationship

The patient-surgeon relationship in emergency surgery is already fragile, often formed minutes before incision. Autonomy can either stabilize trust through reliability and standardization or erode it through perceived delegation and blurred responsibility^[51]. Ethical communication should be proportional to autonomy level and realistic for emergency contexts: patients or surrogates should hear, in plain language, what the system will do, what the surgeon will do, how safety and takeover work, and why this approach is appropriate for the case.

REGULATORY VALIDATION AND PROFESSIONAL STANDARDS

Existing FDA pathways for surgical robots and Software as a Medical Device require bench testing, clinical validation, and post-market surveillance, but were designed for deterministic, bounded systems. They were not built for agents that learn, adapt, and make sequential decisions with patient-safety consequences. This reflects a broader challenge for cyber-physical systems that couple computation to safety-critical physical processes: timing predictability and formal verification, well established for simpler embedded controllers, become substantially harder as systems incorporate learned, adaptive components^[52]. The premarket approval pathway requires substantial clinical evidence and is resource-intensive, while 510(k) clearance depends on identifying a legally marketed predicate device, which does not yet exist for autonomous surgical agents, whereas the De Novo pathway, used for example to authorize the AquaBeam system, is available for novel devices without a predicate but requires the sponsor to establish appropriate risk-based special controls^[53].

The FDA's finalized guidance on Predetermined Change Control Plans (PCCPs) for AI-enabled device software functions is the most relevant existing mechanism for devices expected to change after clearance^[54]. A PCCP allows a manufacturer to pre-specify, and obtain premarket authorization for, a defined envelope of future model modifications together with the methods used to validate them, so that certain updates can be implemented without a new marketing submission. For learning ACS systems, this offers a plausible path to keeping pace with model improvement without repeated *de novo* review; its current limit is that it presumes

a definable modification envelope, which is difficult to specify for an open-ended autonomous agent whose behavior may shift in ways not anticipated at clearance. An early empirical analysis found that public reporting of PCCPs among cleared AI-enabled devices has so far been limited, consistent with this limitation^[55]. [Table 2](#)'s Regulation domain already recommends ACS-specific predetermined change-control plans as one path forward here.

To our knowledge, no society-endorsed standard specific to autonomous operative execution was identified as of the search date. SAGES has published a structured tool for assessing robotic surgery systems generally^[56]; extending it, or a comparable framework from another professional body, to autonomous execution specifically remains an open task.

ACS poses regulatory challenges beyond those facing elective surgical robotics. Emergency consent paradigms (21 CFR 50.24, exception from informed consent for emergency research)^[57] may apply to some clinical trials of autonomous systems in trauma and emergency surgery, yet these pathways are complex and rarely used for device studies; their applicability is context-dependent and should be determined case by case; no default should be assumed. Post-market surveillance must account for the heterogeneity of ACS patients and procedures; a system validated on uncomplicated hernia repairs may perform differently in the hostile abdomen of a patient with perforated diverticulitis and septic shock. Regulators should consider requiring subgroup performance reporting (by disease severity, patient comorbidity, and operative complexity) as a condition of clearance for ACS autonomy applications.

BRIDGING THE TRANSLATIONAL GAPS FOR AUTONOMY IN ACS

[Table 2](#) maps six critical translational gaps: data, validation, uncertainty quantification, fallback protocols, regulation, and consent. Concrete paths forward are proposed for each domain.

EQUITY AND ACCESS

ACS serves diverse, often socioeconomically vulnerable populations. Trauma disproportionately affects young men, communities of color, and uninsured patients. Emergency general surgery is concentrated among the elderly and those with limited access to elective care. If AI systems are trained on non-representative data or deployed only in resource-rich centers, they risk widening the very disparities they might otherwise help address. Equity must therefore be treated as an implementation requirement, operationalized through external validation across sites and subgroup performance reporting after deployment^[17].

The computational burden of autonomous systems intersects with global surgery realities that deserve explicit attention. High-performance inference hardware, real-time imaging pipelines, and continuous connectivity requirements are prerequisites for the systems described in this review. Most low- and middle-income settings, where ACS burden is highest, lack this infrastructure. Lightweight model architectures, edge inference, and tiered deployment frameworks should be built in from the outset, never retrofitted, if ACS autonomy is to reach the patients who need it most.

There is an equity dimension to the pace of deployment as well. If the earliest autonomous tools require expensive robotic platforms and high-bandwidth computing, they will debut in well-resourced academic centers while the rural and low-resource settings that would benefit most from skill equalization wait years or decades longer. A phased capability model may help structure equitable deployment: Tier 1 centers with Graphics Processing Unit (GPU) clusters and full robotic platforms could serve as development and validation hubs; Tier 2 settings with modest edge-computing hardware could run computationally efficient

models for bounded subtasks (camera control, exposure assistance); and Tier 3 environments could access autonomy through tele-supervised architectures, where a remote surgeon oversees an autonomous assistant operating at a distant site, provided that networking latency and safety provisions meet validated thresholds. Tele-supervised autonomy depends on reliable connectivity, local rescue capacity, device maintenance, and the immediate availability of a clinician able to manage system failure; absent these safeguards, it could introduce new vulnerabilities and widen the disparities it was meant to reduce. Federated learning across trauma centers, using privacy-preserving analytics, could enable model improvement from diverse ACS populations without centralizing sensitive patient data. Designing for computational efficiency and platform portability from the outset is an ethical imperative.

LIMITATIONS

This review has several limitations. It is a narrative review: sources were identified through targeted searching and hand-searching of reference lists, without dual independent screening or Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-style reporting, and relevant work may have been missed. The task-autonomy readiness assessment in [Table 1](#) and the translational-gap analysis in [Table 2](#) reflect the expert opinion of a single multidisciplinary author group; they did not undergo a formal Delphi or consensus process, and a broader panel could reach different target autonomy levels, failure-mode characterizations, or fallback behaviors; these tables represent this review's principal original contribution. Finally, this is a fast-moving field: our search extended through preprints available as of 5 March 2026, and specific quantitative claims, particularly those drawn from preprints, may be superseded quickly.

CONCLUSION

ACS may be both the proving ground for surgical autonomy and its ultimate test. It has generated several notable preclinical demonstrations of surgical autonomy to date. Yet the operative conditions that define emergency make it among the most challenging environments in which to deploy these systems safely in human patients. The safest path forward is incremental. Supervised task autonomy first, in bounded domains where safety properties are strong and errors are recoverable. Scope would expand only as staged validation demonstrates reliability under ACS-specific stress conditions. The work ahead extends beyond building better models to building better systems. We must align incentives, stage evaluations, and build governance frameworks and clinical integration pathways that protect patients while enabling responsible innovation. ACS is therefore a demanding test case for surgical autonomy. If bounded autonomy can be made safe in the most unstable, time-critical operative environments in surgery, it would provide a powerful proof-of-concept and a robust safety template that could inform deployment in less demanding settings. Whether and how autonomy will be adopted in ACS remains scientifically and ethically uncertain. What is clear is that its value will depend on whether our evaluation and governance infrastructure matures alongside the technology.

DECLARATIONS

Authors' contributions

Conceptualization: Grunhut J, Jopling JK, Haworth J, Brat G, Krieger A

Literature search and evidence synthesis: Grunhut J, Iyer R, Jopling JK

Writing - original draft: Grunhut J, Iyer R

Writing - review and editing: Grunhut J, Iyer R, Haworth J, Brat G, Krieger A, Jopling JK

Supervision: Jopling JK

All authors read and approved the final manuscript.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tools OpenEvidence (version Osler, released 2026-09-03) and Claude Opus (version 4.8, released 2026-05-28) were used to assist with literature search, citation

verification, and drafting during manuscript revision. The Graphical Abstract was generated with AI tool Claude Opus 4.8. The tools did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg*. 2020;155:148-58. DOI PubMed PMC
2. Arjomandi Rad A, Vardanyan R, Athanasiou T, Maessen J, Sardari Nia P. The ethical considerations of integrating artificial intelligence into surgery: a review. *Interdiscip Cardiovasc Thorac Surg*. 2025;40:ivae192. DOI PubMed PMC
3. Gumbs AA, Frigerio I, Spolverato G, et al. Artificial intelligence surgery: How do we get to autonomous actions in surgery? *Sensors*. 2021;21:5526. DOI PubMed PMC
4. Lee A, Baker TS, Bederson JB, Rapoport BI. Levels of autonomy in FDA-cleared surgical robots: a systematic review. *NPJ Digit Med*. 2024;7:103. DOI PubMed PMC
5. Saeidi H, Opfermann JD, Kam M, et al. Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Sci Robot*. 2022;7:eabj2908. DOI PubMed PMC
6. Kim JWB, Chen JT, Hansen P, et al. SRT-H: a hierarchical framework for autonomous surgery via language-conditioned imitation learning. *Sci Robot*. 2025;10:eadt5254. DOI PubMed
7. Haworth J, Chen JT, Nelson N, et al. SutureBot: a precision framework & benchmark for autonomous end-to-end suturing. *NeurIPS*. 2025. Available from: <https://neurips.cc/virtual/2025/poster/121650>. [Last accessed on 13 Oct 2026].
8. Parasuraman R, Sheridan TB, Wickens CD; New Collective Author. A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern A Syst Hum*. 2000;30:286-97. DOI PubMed
9. Haidegger T. Autonomy for surgical robots: concepts and paradigms. *IEEE Trans Med Robot Bionics*. 2019;1:65-76. DOI
10. Yang GZ, Cambias J, Cleary K, et al. Medical robotics - regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci Robot*. 2017;2:eaam8638. DOI PubMed
11. Opfermann JD, Schmidgall S, Krieger A. State-of-art and the future of autonomous surgery. In: Hung AJ, Editors. Artificial Intelligence in Urology. 2025; pp. 249-73. DOI
12. U. S. Food and Drug Administration. List of artificial intelligence-enabled medical devices. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>. [Last accessed on 13 Oct 2026].
13. Loftus TJ, Altieri MS, Balch JA, et al. Artificial intelligence-enabled decision support in surgery: state-of-the-art and future directions. *Ann Surg*. 2023;278:51-8. DOI PubMed PMC
14. Reinke A, Li ZO, Tizabi MD, et al. Current validation practice undermines surgical AI development. *arXiv* 2025;arXiv:2511.03769. Available from <https://doi.org/10.48550/arXiv.2511.03769> [accessed 13 October 2026].
15. Limon D, Satish V, Raghavan N, Nguyen P, Rajesh A. Artificial intelligence in surgery revisited: a 2025 guide to understanding and applying AI models in clinical practice. *Am Surg*. 2026;92:687-97. DOI PubMed
16. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70-6. DOI PubMed PMC

17. De Simone B, Deeken G, Catena F. Balancing ethics and innovation: Can artificial intelligence safely transform emergency surgery? A narrative perspective. *J Clin Med.* 2025;14:3111. DOI PubMed PMC
18. Marcus HJ, Ramirez PT, Khan DZ, et al. ; Additional collaborators. The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring. *Nat Med.* 2024;30:61-75. DOI PubMed PMC
19. Brattain LJ, Pierce TT, Gjestebj LA, et al. AI-enabled, ultrasound-guided handheld robotic device for femoral vascular access. *Biosensors.* 2021;11:522. DOI PubMed PMC
20. Herlihy JP, Cohn WE, Ebner A. Centrally inserted central catheter placement using a novel, handheld, image-guided, robotic device: Results of initial feasibility trial in patients. *J Vasc Access.* 2025;26:2027-34. DOI PubMed PMC
21. Grunhut J, Walker P, Lackie M, Patterson K, Driscoll K, Sarkar R. Autonomous vascular access devices: current solutions and future challenges. *Mil Med.* 2026:usag377. DOI PubMed
22. Schmidgall S, Opfermann JD, Kim JW, Krieger A. Will your next surgeon be a robot? Autonomy and AI in robotic surgery. *Sci Robot.* 2025;10:eadt0187. DOI PubMed PMC
23. Zhang R, Chen J, Wang Z, et al. A step towards conditional autonomy - robotic appendectomy. *IEEE Robot Autom Lett.* 2023;8:2429-36. DOI
24. Schüßler A, Kunz C, Younis R, et al. Semi-autonomous robotic assistance for gallbladder retraction in surgery. *IEEE Robot Autom Lett.* 2025;10:7468-75. DOI
25. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med.* 2016;8:337ra64. DOI PubMed
26. Ward TM, Mascagni P, Ban Y, et al. Computer vision in surgery. *Surgery.* 2021;169:1253-6. DOI PubMed
27. Long Y, Lin A, Kwok DHC, et al. Surgical embodied intelligence for generalized task autonomy in laparoscopic robot-assisted surgery. *Sci Robot.* 2025;10:eadt3093. DOI PubMed
28. Schmidgall S, Kim JW, Kuntz A, Ghazi AE, Krieger A. General-purpose foundation models for increased autonomy in robot-assisted surgery. *Nat Mach Intell.* 2024;6:1275-83. DOI
29. Liu T, Yuan S, Ren H. Artificial kinaesthesia in autonomous robotic surgery. *Nat Rev Bioeng.* 2026;4:210-1. DOI
30. Li Y, Richter F, Lu J, et al. SuPer: a surgical perception framework for endoscopic tissue manipulation with surgical robotics. *IEEE Robot Autom Lett.* 2020;5:2294-301. DOI
31. Lin S, Miao AJ, Alabiad A, et al. SuPerPM: a surgical perception framework based on deep point matching learned from physical constrained simulation data. *arXiv* 2025;arXiv:2309.13863. Available from <https://doi.org/10.48550/arXiv.2309.13863> [accessed 13 October 2026].
32. Liang X, Liu F, Zhang Y, Li Y, Lin S, Yip M. Real-to-sim deformable object manipulation: optimizing physics models with residual mappings for robotic surgery. *arXiv* 2024;arXiv:2309.11656. Available from <https://doi.org/10.48550/arXiv.2309.11656> [accessed 13 October 2026].
33. Ge J, Kilmer E, Mady LJ, Opfermann JD, Krieger A. Enhancing surgical precision in autonomous robotic incisions via physics-based tissue cutting simulation. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2024 Oct 14-18; Abu Dhabi, United Arab Emirates. IEEE; 2024. pp. 2421-8. DOI PubMed PMC
34. Ou Y, Tavakoli M. Sim-to-real surgical robot learning and autonomous planning for internal tissue points manipulation using reinforcement learning. *IEEE Robot Autom Lett.* 2023;8:2502-9. DOI
35. Yu Q, Moghani M, Dharmarajan K, et al. Orbit-surgical: an open-simulation framework for learning surgical augmented dexterity. *arXiv* 2024;arXiv:2404.16027. Available from <https://doi.org/10.48550/arXiv.2404.16027> [accessed 13 October 2026].
36. Ho YJ, Chiu ZY, Zhi Y, Yip MC. SurgIRL: towards life-long learning for surgical automation by incremental reinforcement learning. *IEEE Robot Autom Lett.* 2025;10:13145-52. DOI
37. Liang X, Wang CP, Shinde NU, Liu F, Richter F, Yip M. MEDiC: autonomous surgical robotic assistance to maximizing exposure for dissection and cautery. *arXiv* 2024;arXiv:2409.14287. Available from <https://doi.org/10.48550/arXiv.2409.14287> [accessed 13 October 2026].
38. Wang J, Yao M, Wei Y, Guo X, Zheng A, Zhao W. Image-to-force estimation for soft tissue interaction in robotic-assisted surgery using structured light. *IEEE Robot Autom Lett.* 2025;10:7795-802. DOI
39. Reyzaal MDI, Chen M, Huang W, Ourselin S, Liu H. DaFoEs: mixing datasets towards the generalization of vision-state deep-learning force estimation in minimally invasive robotic surgery. *IEEE Robot Autom Lett.* 2024;9:2527-34. DOI
40. Mazidi A, Roshanfar M, Sayadi A, et al. Learning-based force sensing and impedance matching for safe haptic feedback in robot-assisted laparoscopic surgery. *arXiv* 2026;arXiv:2601.14445. Available from <https://doi.org/10.48550/arXiv.2601.14445> [accessed 13 October 2026].
41. Liu T, Zhang T, Katupitiya J, Wang J, Wu L. Haptics-enabled forceps with multimodal force sensing: toward task-autonomous surgery. *IEEE/ASME Trans Mechatron.* 2024;29:2208-19. DOI

42. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D. Concrete problems in AI safety. *arXiv* 2016;arXiv:1606.06565. Available from <https://doi.org/10.48550/arXiv.1606.06565> [accessed 13 October 2026].
43. Vasey B, Lippert KA, Khan DZ, et al. Intraoperative applications of artificial intelligence in robotic surgery: a scoping review of current development stages and levels of autonomy. *Ann Surg*. 2023;278:896-903. DOI PubMed PMC
44. Guzman NM, Marwaha JS, Keller DS, et al. Regulating digital surgery to balance safety and innovation: a SAGES white paper. *Surg Endosc*. 2026;40:5480-90. DOI PubMed PMC
45. Kelkar D, Borse MA, Godbole GP, Kurlekar U, Slack M. Interim safety analysis of the first-in-human clinical trial of the Versius surgical system, a new robot-assisted device for use in minimal access surgery. *Surg Endosc*. 2020;35:5193-202. DOI PubMed PMC
46. Fraser AG, Buccheri S, Byrne RA, et al. Recommended methodologies for clinical investigations of high-risk medical devices - conclusions from the European Union CORE-MD Project. *Lancet Reg Health Eur*. 2025;58:101460. DOI PubMed PMC
47. Xu S, Perez M, Yang K, Perrenot C, Felblinger J, Hubert J. Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer® simulator. *Surg Endosc*. 2014;28:2569-76. DOI PubMed
48. Richter F, Shen S, Liu F, et al. Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection. *IEEE Robot Autom Lett*. 2021;6:1383-90. DOI
49. Liang Z, Liang X, Atar S, et al. LapSurgie: humanoid robots performing surgery via teleoperated handheld laparoscopy. *arXiv* 2026;arXiv:2510.03529. Available from <https://doi.org/10.48550/arXiv.2510.03529> [accessed 13 October 2026].
50. Hashimoto DA, Marwaha JS, Lee SA, Schwaitzberg S, Duffourc MN. Risk and liability in the deployment of AI systems for surgery: a SAGES white paper. *Surg Endosc*. 2026;40:5468-79. DOI PubMed PMC
51. Subramanian T, Pocivavsek L, Alverdy JC. To operate or not? Balancing advanced imaging, machine learning, and the doctor-patient relationship in complex clinical decision making. *Acad Med*. 2023;98:S34-6. DOI PubMed PMC
52. Lee EA. Cyber physical systems: design challenges. UC Berkeley EECS Department Technical Report No. UCB/EECS-2008-8; 2008. Available from: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-8.html>. [Last accessed on 13 Oct 2026].
53. Aboy M, Crespo C, Stern A. Beyond the 510(k): the regulation of novel moderate-risk medical devices, intellectual property considerations, and innovation incentives in the FDA's De Novo pathway. *NPJ Digit Med*. 2024;7:29. DOI PubMed PMC
54. U. S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. Guidance for Industry and Food and Drug Administration Staff. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence>. [Last accessed on 13 Oct 2026].
55. Dabek PA, Bourgeois FT, Stern AD, Zhu Y, Everhart AO. Use and public reporting of predetermined change control plans for AI-enabled medical devices. *JAMA Health Forum*. 2026;7:e261204. DOI PubMed PMC
56. Cha JS, Fuller P, Ball M, Samreen S, Sylla P, Stefanidis D. ; SAGES Robotic Committee. SAGES tool for assessing robotic surgery systems (STARSS). *Surg Endosc*. 2025;39:5479-91. DOI PubMed PMC
57. Electronic Code of Federal Regulations. 21 CFR 50.24 - Exception from informed consent requirements for emergency research. Available from: <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-50/subpart-B/section-50.24>. [Last accessed on 13 Oct 2026].
58. Gruijthuijsen C, Garcia-Peraza-Herrera LC, Borghesan G, et al. Robotic endoscope control via autonomous instrument tracking. *Front Robot AI*. 2022;9:832208. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.