**Regulating Digital Surgery to Balance Safety and Innovation**

*Natalie M. Guzman, Jayson S. Marwaha… [additional authors] Binita S. Ashar, Amin Madani, Daniel A. Hashimoto, Dana A. Telem*

Word count: ~3,500

## INTRODUCTION

Digital surgery represents a paradigm shift in surgical practice that inserts a computer interface between the surgeon and patient, reshaping how surgical procedures are performed, monitored, and assessed.[1] Robotic surgical systems can enable minimally invasive procedures with enhanced dexterity and control, while artificial intelligence algorithms will be able to provide real-time analysis of surgical video feeds to identify critical anatomy and potential complications.[2,3] Similarly, augmented reality (AR) systems can overlay digital information onto the surgical field, helping to guide instrument placement and improve navigation accuracy.[4–6] Additionally, machine learning (ML) models can enable the analysis of large datasets of surgical outcomes to predict patient risk and optimize treatment plans.[7–9]

While these technologies have tremendous potential to transform surgical care, they also introduce significant risks. Artificial Intelligence (AI) algorithms may provide incorrect recommendations during critical procedures, AR systems may display inaccurate visualizations that mislead surgeons, and robotic systems can malfunction during operations. Machine learning models may exhibit bias against certain patient populations, and autonomous surgical systems could make life-threatening errors without human oversight.[10,11]

AI/ML, when incorporated into surgical and procedural technologies, carries distinct risks because unlike diagnostic tools, they guide real-time, irreversible actions, where there is no time to cross-check and correct incorrect recommendations. Additionally, they often learn and adapt over time, integrate multiple hardware and software components, process real-time data streams, and require the characterization of metrics for measuring performance and safety.[12] As these systems become more sophisticated, they will require

equally sophisticated regulatory approaches capable of ensuring thorough evaluation of high-risk technologies while avoiding unnecessarily burdensome requirements for low-risk tools.

In this paper, we characterize the unique regulatory challenges of digital surgery technologies. Following this discussion, we provide guidance on general regulatory principles specific to digital surgery technologies that enable safe implementation while allowing for efficient, expedited delivery of these beneficial technologies to patients.

## DIGITAL SURGERY REGULATORY CHALLENGES

### Enhanced Instrumentation and Robotics

Enhanced surgical instruments and robotics represent a fundamental transformation in medical technology. These technologies include powered staplers with tissue compression sensing, advanced energy devices with impedance-based feedback, and roboticized devices that integrate robotic technology into traditional surgical instruments.[1] Until recently, surgical tools have been static instruments that simply extend the surgeon's capabilities, but we now have access to intelligent systems equipped with sensors, automation capabilities, and varying levels of autonomy.[1,13] This evolution has created unprecedented regulatory challenges, as current device classification systems were designed for medical devices with fixed functionality rather than hybrid hardware-software systems that integrate artificial intelligence technology.[14] The United States Food and Drug Administration (FDA) is moving towards addressing these challenges and recently issued final guidance providing recommendations for predetermined change control plans (PCCPs) tailored to AI-enabled devices.[15] Yet, significant challenges persist.

The lack of clinically relevant and validated performance metrics represents a critical regulatory gap for enhanced instrumentation. While various performance metrics exist for these adaptive technologies, current regulatory frameworks often rely on technical benchmarks that may not translate to meaningful clinical outcomes or patient safety measures, making it difficult to assess real-world safety and efficacy consistently across different devices and clinical applications. This challenge is compounded by evidence that

validation metrics in AI healthcare technology are often inadequately chosen and poorly correlated with clinical outcomes, which can hinder the translation of these tools into clinical practice.[16] Without metrics grounded in outcome data, regulators cannot effectively measure device performance, compare competing technologies, or establish meaningful safety thresholds for approval.[17] Performance metrics must account for both the mechanical components and the software algorithms that control the device's actions, requiring evaluation methodologies that can assess integrated systems in a sophisticated manner.

While the FDA's PCCP guidance for AI-enabled device software functions represents an important step forward in addressing software modifications and the continuous learning nature of these systems, there is still ambiguity to be addressed. When there is collaboration between device manufacturers and software developers in creating these hybrid systems, this leads to ambiguous liability scenarios that current regulations fail to address, leaving unclear responsibilities for device performance, safety monitoring, and adverse event reporting across the various stakeholders involved in these integrated technologies.[18,19]

**Advanced Visualization Technologies**

Advanced visualization systems build upon the challenges of enhanced instrumentation while introducing unique complexities around real-time image processing, registration accuracy, human factors considerations, and cybersecurity.[1,20–24] These technologies include three-dimensional visualization, fluorescence-guided surgery, and AR systems that provide real-time visual information overlaid on the surgical field.[1] Similarly to enhanced instrumentations and robotics, the adaptive nature of these technologies challenges traditional performance metrics.[25] Visualization systems rely on real-time image processing algorithms that must continuously adapt to changing surgical conditions, varying patient anatomy, and different lighting environments. This creates uncertainty about establishing appropriate safety thresholds for technologies whose output varies based on contextual factors that cannot be fully predicted during pre-market testing.[26]

Registration accuracy represents one of the most critical performance metrics, particularly for AR applications where virtual objects must align precisely with real anatomy.[27] Registration failures can lead directly to surgical errors with potentially catastrophic outcomes.[28] For example, in cardiac surgery, registration errors as small as 0.2 mm with coronary vessels could damage critical structures, cause chamber perforation, result in incorrect device placement, and potentially lead to death. Current regulatory frameworks lack standardized metrics for evaluating registration performance, and this challenge is compounded by the absence of established quality metrics for registration evaluation itself.

These systems also present unique challenges for clinical trial methodologies. Unlike traditional devices evaluated with standardized protocols, visualization systems must be assessed within their intended clinical environments where real-world factors, such as a surgeon's skill and or the real data source, significantly influence performance.[29] This necessitates evaluation methodologies extending beyond traditional randomized controlled trials to include comprehensive implementation studies and real-world evidence collection.[30]

Human factors considerations present another regulatory gap, as advanced visualization systems directly impact surgeon cognitive load, visual attention, and decision-making processes.[21,23,24,31] Even minor side effects like fatigue, disorientation, or visual obstruction could become dangerous during critical surgical moments.[32]

Additionally, cybersecurity concerns arise from the networked infrastructure required for image processing and data storage, resulting in potential exposures that could compromise patient safety.[33] Current regulatory frameworks could be improved with comprehensive cybersecurity standards specifically tailored to real-time surgical visualization systems, where security breaches could have immediate and severe consequences during active procedures.

**AI/ML Data Analytics and Data Capture**

AI and ML systems integrated with comprehensive data analytics and capture represent one of the most complex regulatory challenges in digital surgery. This encompasses real-time surgical decision support algorithms, predictive analytics for patient outcomes, automated surgical video analysis, and extensive perioperative data collection platforms that capture everything from instrument kinematics to surgeon performance metrics.[1] In addition to some of the factors already discussed, AI/ML systems pose multiple regulatory challenges.

A known reality of these models is performance drift. Unlike traditional devices where performance degradation typically results from mechanical wear or component failure, AI/ML systems experience performance drift when real-world data differs from training data, when patient populations change, or when clinical workflows evolve.[34–37] Current regulatory frameworks lack adequate direction on mechanisms for detecting, monitoring, and responding to such performance drift.

Again, the issue of meaningful metrics arises. It is unclear if developers should be prioritizing algorithmic performance measures like sensitivity and specificity or clinical performance indicators like surgical outcomes and patient safety, or more realistically, a balance of both. The choice of evaluation metrics fundamentally affects how these systems are assessed, approved, and monitored, yet a standardized approach for determining which metrics are appropriate for different AI/ML applications in digital surgery does not exist. This ambiguity may lead to approval of systems that perform well on technical metrics but fail to improve clinical outcomes, compromising patient safety.

Algorithmic bias presents another important regulatory gap. AI/ML systems can perpetuate or amplify existing healthcare disparities when training data lacks diversity or when algorithms are optimized for majority populations.[38,39] In surgical applications, this could mean that decision support systems perform well for certain demographic groups while providing suboptimal recommendations for others, potentially exacerbating existing inequities in surgical care.

Data capture and ownership issues also challenge traditional regulatory frameworks. Digital surgery platforms can generate unprecedented amounts of data, including surgical video recordings, instrument kinematics, audio recordings, and comprehensive procedure documentation.[1] This data has immense value for quality improvement, surgeon training, and algorithm development, but there are still many questions related to data governance, ownership rights, privacy protection, and appropriate use limitations.[40]

## GENERAL PRINCIPLES FOR REGULATING DIGITAL SURGERY

### Existing Regulatory Landscape of Digital Surgery

Digital surgery and procedural technologies have unique risk considerations that differ from other digital tools in medicine. Procedural technologies provide guidance for irreversible actions, unlike diagnostic decision support systems where it is more feasible that incorrect recommendations could be identified and corrected. For example, if a surgeon had technology that incorrectly guides them to cut the bile duct or misidentifies critical anatomy during a cardiac procedure, the consequences are immediate and, in many cases, life-threatening. This procedural versus non-procedural distinction require regulatory approaches more aligned with current medical device requirements than traditional software evaluation, which is the path we are seeing.

The FDA currently regulates digital surgery technologies through 510(k) clearance for substantially equivalent devices, De Novo classification for novel low-to-moderate risk devices, and premarket approval (PMA) for high-risk devices.[34,41] Current examples demonstrate how different types of digital surgery technologies have navigated these pathways successfully. For instance, navigation systems like Proprio's Paradigm AI Surgical Guidance Platform and robotic systems like Moon Surgical's Maestro System with AI-powered ScoPilot have received 510(k) clearance, demonstrating substantial equivalence to existing technologies.[42,43] Meanwhile, novel robotic systems without clear predicates, such as Virtual Incision's MIRA Surgical System for miniaturized robotic-assisted surgery and CMR Surgical's Versius Surgical Robotic System, have successfully navigated the De

Novo pathway.[44,45] The Premarket Approval (PMA) pathway remains reserved for the highest-risk innovations, as exemplified by Perimeter Medical Imaging's B-Series OCT with ImgAssist AI 2.0, which is currently under FDA review for real-time surgical margin assessment during breast-conserving surgeries.[46] Currently, PMA is rare for AI surgical tools.

Given the specific regulatory challenges with digital surgery tools and technologies, it is important to consider how these systems require different regulatory approaches than traditional medical devices. While these FDA pathways provide a foundation, they inadequately address some of the unique complexities specific to digital surgery and procedural technologies. The America's AI Action Plan, the FDA's AI/ML Software as a Medical Device Action Plan, and PCCPs represent important steps forward, but we will have to further tailor our regulatory guidance to ensure efficient and safe regulations of these tools.[41,47–49] Building on the FDA's existing framework, we propose a regulatory structure that incorporates enhanced risk-based principles specifically designed for the procedural context of digital surgery and to account for the fact that errors in these applications can carry immediate and irreversible patient harm (**Table 1**).

**Three Tier Risk Classification System**

*Low Risk Technologies*

Low risk technologies include digital surgery technologies that pose minimal potential for direct patient harm (**Table 1**). This includes training simulators, basic surgical planning software without real-time guidance, and data analytics platforms for quality improvement. Examples include AI-enabled imaging analysis tools such as Aidoc's BriefCase, which screens CT scans for incidental findings, and Viz Vascular Suite, which detects vascular pathologies to aid in workflow prioritization.[50] These technologies should undergo streamlined regulatory pathways that have reduced documentation requirements but still should include mandatory post-market surveillance. The regulatory burden should focus on basic safety validation, cybersecurity standards, and user interface design factors.

*Moderate Risk Technologies*

Moderate risk technologies include real-time surgical guidance systems, advanced visualization tools with decision support capabilities, and semi-autonomous instruments that require human oversight (**Table 1**). Examples include Proprio's Paradigm platform, which provides AI-guided visualization and real-time intraoperative measurements for spine surgery, and Moon Surgical's Maestro System with ScoPilot, which enables AI-powered laparoscope control during minimally invasive procedures.[42,43] These technologies should require more comprehensive pre-market evaluation, including clinical validation studies, human factors assessment, and algorithmic bias evaluation. There is a need for continuous performance monitoring with predefined intervention thresholds, and many of these technologies should include PCCPs. There should also be the requirement of evidence of safe integration within existing surgical workflows.

*High Risk Technologies*

High risk technologies include autonomous or semi-autonomous surgical systems, AI-guided critical anatomical navigation, and any technology that can independently initiate irreversible actions (**Table 1**). Examples include novel robotic systems like the MIRA Surgical System, which represents the first miniaturized robotic-assisted surgery device cleared through De Novo for colectomy procedures, and CMR Surgical's Versius system for robotic-assisted gallbladder removal.[44,45] The highest-risk category includes systems like Perimeter's B-Series OCT with ImgAssist AI 2.0, which provides real-time assessment of surgical margins during breast cancer surgery and is currently under review via the PMA pathway due to its novel application and direct impact on immediate surgical decision-making.[46] These require the most stringent level of regulation through either De Novo or PMA pathways with comprehensive safety and risk mitigation protocols in addition to rigorous randomized clinical trials. For these trials, there should also be hybrid effectiveness-implementation designs as AI interventions' effectiveness is often context-specific, making implementation measures one in the same with effectiveness measures.[30]

*Predetermined Change Control Plans*

A critical component of AI regulation that is relevant regardless of risk levels, are PCCPs. PCCPs represent a proactive regulatory approach that addresses the adaptive nature of AI-enabled devices by establishing pre-authorized pathways for modifications that would otherwise require additional FDA submissions.[15] For companies developing digital surgery technologies that incorporate continuous learning algorithms or anticipate iterative improvements, PCCPs provide regulatory efficiency without compromising patient safety oversight.

These plans must specify the Description of Modifications that can be made, establish a detailed Modification Protocol with predefined acceptance criteria, and include a comprehensive Impact Assessment that evaluates benefits and risks.[15] Particularly for moderate and high-risk digital surgery technologies, PCCPs enable manufacturers to implement algorithm refinements, performance optimizations, and safety enhancements in response to real-world data while maintaining continuous regulatory compliance. Without well-designed PCCPs, the iterative nature of AI-enabled surgical technologies could create significant regulatory bottlenecks that delay beneficial improvements from reaching patients, particularly as these systems learn and adapt from clinical use and new data points.

**Performance Metrics and Measurement Standards**

One of the most significant barriers to effective digital surgery regulation is having standardized performance metrics that are clinically meaningful. Evaluation approaches that prioritize inappropriately defined technical metrics like sensitivity and specificity over clinical outcomes can lead to approval of systems that perform well algorithmically but may fail to improve patient care. Furthermore, performance metrics frequently fail to reflect what matters most in the specific clinical application, inadequately measuring meaningful scientific progress and limiting practical translation of AI/ML techniques.[16,51] Regulatory frameworks must balance evaluation criteria that assess both technical performance and clinical impact.

Technical performance standards should include accuracy metrics specific to each technology type, such as registration precision for AR systems, response time requirements for real-time guidance systems, and failure detection capabilities for autonomous components.[52] These standards must be validated across diverse patient populations and clinical environments to ensure generalizability and identify potential algorithmic bias.[53]

Clinical performance standards should demonstrate meaningful improvement in surgical outcomes, workflow efficiency, or patient safety. This requires moving beyond traditional randomized controlled trials to include pragmatic trial designs that capture real-world implementation challenges, adaptive study protocols that account for learning algorithm evolution, and long-term outcome tracking that extends beyond immediate procedural success.[30,54]

Safety monitoring standards should establish continuous surveillance systems that detect performance drift, monitor adverse events, and track user satisfaction.[25] These systems should include automated reporting mechanisms, standardized incident classification systems, and clear escalation procedures when performance thresholds are not met.[52]

*Role of Professional Societies and Collaborative Communities*

Professional societies and collaborative communities could play a crucial role in bridging the gap between technical performance metrics and clinically meaningful evaluation standards. Manufacturers develop metrics to generate evidence for regulatory submissions and the FDA then assesses the appropriateness and rigor of these metrics. This can work well for technical performance standards, but for clinical performance standards and safety monitoring requirements, professional societies such as the American College of Surgeons (ACS) and the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) could provide essential clinical expertise that ensures evaluation metrics reflect real-world surgical priorities and patient safety concerns.

These societies could go a step further to form collaborative communities. These communities are forums where public and private sector members work together to achieve common objectives and solve shared challenges related to medical device development and regulation.[55] Collaborative communities develop best practices and strategies for challenges, generate and evaluate evidence for new approaches, and implement solutions.[55] They also may work to clarify ill-defined challenges or generate consensus on the definition and scope of the challenge which can aid in tailoring appropriate strategies to tackle those challenges.[55] These communities would best be convened by interested stakeholders. For digital surgery regulation, these communities could bring together diverse stakeholders including surgeons, device manufacturers, patients, healthcare system leaders, and potentially FDA representatives to collectively determine and recommend appropriate performance metrics and evaluation standards.

The value of these professional societies and collaborative approaches lies in their ability to identify what metrics matter most in real clinical scenarios. For example, consider an AI-enabled device that detects melanoma.[56] A device with 90% sensitivity and specificity might initially sound inferior to one with 96% sensitivity but only 30% specificity. However, clinical expertise becomes critical here because the device with the higher sensitivity is actually preferable in this clinical scenario due to the fact that missing a melanoma diagnosis can have devastating consequences, often including death. When melanoma is caught while localized, the 5-year survival rate exceeds 99%, but if the cancer spreads to distant parts of the body, the 5-year survival rate plummets to only 35%.[57] However, for detecting other skin lesions where the consequences of missing the diagnosis are much less severe, specificity becomes more important because clinicians want to balance unnecessary biopsies or treatments with identification. This example demonstrates how clinical perspective shapes which performance metrics matter most for patient safety and outcomes, balancing optimal metrics based on disease severity and treatment implications.

While manufacturers may be able to demonstrate impressive technical performance metrics for surgical technologies, surgical societies and collaborative

communities can ensure the evaluation criteria also capture the most critical clinical aspects of procedures, such as confirming that an AI tool accurately identifies the critical view of safety during cholecystectomy or properly recognizes high-risk anatomical landmarks during complex procedures. When these communities can bring diverse stakeholders together to provide recommendations and consensus on important clinical endpoints, the regulatory process for these tools could be streamlined since the FDA values meaningful stakeholder engagement in its assessments.

**Evidence Requirements and Burden of Proof**

The burden of proof for digital surgery technologies must reflect the irreversible nature of surgical procedures and the potential for significant, immediate, and permanent patient harm. This necessitates comprehensive evidence generation before market entry and continued validation throughout the product lifecycle.

Pre-market evidence requirements should continue to be tiered based on risk classification. Low-risk technologies can require basic safety validation and usability testing, while high-risk systems necessitate extensive clinical trials with statistically powered safety endpoints. The European CORE–MD consortium (Coordinating Research and Evidence for Medical Devices) has developed a practical clinical risk score that provides a structured approach to determining appropriate evidence requirements for AI-enabled medical devices, which we could build on.[37]

This scoring system evaluates three key domains. First is a valid clinical association, which considers how transparency and oversight can be achieved, and that there is a clear and valid association of the technology with its targeted indication. Second is valid technical performance, which considers the strength of validation and how rigorously the technology has been tested. Third is clinical performance, which considers the use context (i.e., disease type, condition, healthcare situation) and the function of the output (i.e., inform vs drive vs diagnose or treat).

Higher composite scores indicate devices that require extensive clinical investigations before regulatory approval, while lower scores suggest devices could be

approved with less rigorous pre-market evaluation. This scoring system could be adapted for digital surgery technologies in the United States to ensure that regulatory requirements are proportionate to actual risk. Regardless of the risk tier, the evidence should demonstrate not only that the technology performs as intended in controlled settings, but that it integrates safely with existing surgical workflows, does not introduce increased cognitive burdens for surgeons, and ultimately improves patient outcomes rather than simply performs well with technical metrics.

Post-market evidence generation and additional data collection should also be mandatory for all these technologies,[37] with reporting requirements that scale with risk level. Low-risk technologies should report basic usage patterns and adverse events, while high-risk systems should require comprehensive real-world evidence collection including detailed outcome tracking, performance monitoring, and comparative effectiveness studies. Digital surgery technologies must prove their effectiveness not just in controlled research environments, but they must prove effective in the context of their intended use.[58] This requires evidence collection across multiple institutions, diverse patient populations, and varying levels of user expertise

**Conclusions**

Regulation of digital surgery must balance innovation with safety, recognizing that both excessive and insufficient regulation can harm patients. Overly restrictive approaches delay beneficial technologies and stifle innovation, while inadequate oversight allows dangerous or ineffective systems to reach patients. The risk-stratified approach outlined above provides a foundation for achieving this balance through evidence-based standards, appropriate burden of proof requirements, and evaluation processes that address the unique challenges of digital surgery technologies.

**REFERENCES**


1. Ali JT, Yang G, Green CA, et al. Defining digital surgery: a SAGES white paper. *Surg Endosc*. 2024;38(2):475-487. doi:10.1007/s00464-023-10551-7

2. Baldari L, Boni L, Cassinotti E. Hybrid robotic systems. *Surgery*. 2024;176(5):1538-1541. doi:10.1016/j.surg.2024.07.049

3. Ward TM, Hashimoto DA, Ban Y, Rosman G, Meireles OR. Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation. *Surg Endosc*. 2022;36(9):6832-6840. doi:10.1007/s00464-022-09009-z

4. Gupta A, Ruijters D, Flexman ML. Augmented Reality for Interventional Procedures. In: *Digital Surgery*. Springer, Cham; 2021:233-246. doi:10.1007/978-3-030-49100-0_17

5. Shuhaiber JH. Augmented Reality in Surgery. *Arch Surg*. 2004;139(2):170-174. doi:10.1001/archsurg.139.2.170

6. Sun P, Zhao Y, Men J, et al. Application of Virtual and Augmented Reality Technology in Hip Surgery: Systematic Review. *J Med Internet Res*. 2023;25:e37599. doi:10.2196/37599

7. Li YY, Wang JJ, Huang SH, et al. Implementation of a machine learning application in preoperative risk assessment for hip repair surgery. *BMC Anesthesiol*. 2022;22:116. doi:10.1186/s12871-022-01648-y

8. Elfanagely O, Toyoda Y, Othman S, et al. Machine Learning and Surgical Outcomes Prediction: A Systematic Review. *Journal of Surgical Research*. 2021;264:346-361. doi:10.1016/j.jss.2021.02.045

9. Xue B, Li D, Lu C, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. *JAMA Netw Open*. 2021;4(3):e212240. doi:10.1001/jamanetworkopen.2021.2240

10. Adegbesan A, Akingbola A, Aremu O, Adewole O, Amamdikwa JC, Shagaya U. From Scalpels to Algorithms: The Risk of Dependence on Artificial Intelligence in Surgery. *Journal of Medicine, Surgery, and Public Health*. 2024;3:100140. doi:10.1016/j.glmedi.2024.100140

11. Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. *Nat Med*. 2024;30(5):1257-1268. doi:10.1038/s41591-024-02970-3

12. Zhou K, Gattinger G. The Evolving Regulatory Paradigm of AI in MedTech: A Review of Perspectives and Where We Are Today. *Ther Innov Regul Sci*. 2024;58(3):456-464. doi:10.1007/s43441-024-00628-3

13. Dario P, Hannaford B, Menciassi A. Smart surgical tools and augmenting devices. *IEEE Transactions on Robotics and Automation*. 2003;19(5):782-792. doi:10.1109/TRA.2003.817071

14. Health C for D and R. Artificial Intelligence in Software as a Medical Device. *FDA*. Published online July 10, 2025. Accessed August 13, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device

15. Center for Devices and Radiological. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions. August 18, 2025. Accessed September 8, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence

16. Reinke A, Tizabi MD, Baumgartner M, et al. Understanding metric-related pitfalls in image analysis validation. *Nat Methods*. 2024;21(2):182-194. doi:10.1038/s41592-023-02150-0

17. O'Sullivan S, Nevejans N, Allen C, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robot*. 2019;15(1):e1968. doi:10.1002/rcs.1968

18. The Yale Review Of International Studies. Navigating Liability In Autonomous Robots: Legal And Ethical Challenges In Manufacturing And Military Applications. March 6, 2025. Accessed August 13, 2025. https://yris.yira.org/column/navigating-liability-in-autonomous-robots-legal-and-ethical-challenges-in-manufacturing-and-military-applications/

19. Shentu X. A review on legal issues of medical robots. *Medicine (Baltimore)*. 2024;103(21):e38330. doi:10.1097/MD.0000000000038330

20. Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. *Med Image Anal*. 2017;37:66-90. doi:10.1016/j.media.2017.01.007

21. Guanà R, Ferrero L, Garofalo S, et al. Skills Comparison in Pediatric Residents Using a 2-Dimensional versus a 3-Dimensional High-Definition Camera in a Pediatric Laparoscopic Simulator. *J Surg Educ*. 2017;74(4):644-649. doi:10.1016/j.jsurg.2016.12.002

22. Pratt P, Stoyanov D, Visentini-Scarzanella M, Yang GZ. Dynamic guidance for robotic surgery using image-constrained biomechanical models. *Med Image Comput Comput Assist Interv*. 2010;13(Pt 1):77-85. doi:10.1007/978-3-642-15705-9_10

23. Condino S, Carbone M, Piazza R, Ferrari M, Ferrari V. Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. *IEEE Trans Biomed Eng*. 2020;67(2):411-419. doi:10.1109/TBME.2019.2914517

24. Edwards PJ“ E, Chand M, Birlo M, Stoyanov D. The Challenge of Augmented Reality in Surgery. In: *Digital Surgery*. Springer, Cham; 2021:121-135. doi:10.1007/978-3-030-49100-0_10

25. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. doi:10.1186/s12916-019-1426-2

26. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. *Science Robotics*. 2017;2(6):eaan6080. doi:10.1126/scirobotics.aan6080

27. Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. *Medical Image Analysis*. 2017;37:66-90. doi:10.1016/j.media.2017.01.007

28. Kos TM, Colombo E, Bartels LW, Robe PA, van Doormaal TPC. Evaluation Metrics for Augmented Reality in Neurosurgical Preoperative Planning, Surgical Navigation, and Surgical Treatment Guidance: A Systematic Review. *Oper Neurosurg (Hagerstown)*. 2024;26(5):491-501. doi:10.1227/ons.0000000000001009

29. Malhotra S, Halabi O, Dakua SP, Padhan J, Paul S, Palliyali W. Augmented Reality in Surgical Navigation: A Review of Evaluation and Validation Metrics. *Applied Sciences*. 2023;13(3):1629. doi:10.3390/app13031629

30. Guzman NM, Kilbourne AM, Telem DA. Applied Artificial Intelligence: The Next Frontier. *Ann Surg*. Published online August 6, 2025. doi:10.1097/SLA.0000000000006879

31. Pratt P, Stoyanov D, Visentini-Scarzanella M, Yang GZ. Dynamic Guidance for Robotic Surgery Using Image-Constrained Biomechanical Models. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. Springer, Berlin, Heidelberg; 2010:77-85. doi:10.1007/978-3-642-15705-9_10

32. FDA. PEAC Executive Summary– Augmented Reality (AR) and Virtual Reality (VR) Medical Devices. Published online July 12, 2022. Accessed June 3, 2025. https://www.fda.gov/medical-devices/digital-health-center-excellence/augmented-reality-and-virtual-reality-medical-devices

33. Gordon WJ, Ikoma N, Lyu H, Jackson GP, Landman A. Protecting procedural care—cybersecurity considerations for robotic surgery. *npj Digit Med*. 2022;5(1):148. doi:10.1038/s41746-022-00693-8

34. Warraich HJ, Tazbaz T, Califf RM. FDA Perspective on the Regulation of Artificial Intelligence in Health Care and Biomedicine. *JAMA*. 2025;333(3):241-247. doi:10.1001/jama.2024.21451

35. Gilbert S, Fenech M, Hirsch M, Upadhyay S, Biasiucci A, Starlinger J. Algorithm Change Protocols in the Regulation of Adaptive Machine Learning-Based Medical Devices. *J Med Internet Res*. 2021;23(10):e30545. doi:10.2196/30545

36. McKee M, Wouters OJ. The Challenges of Regulating Artificial Intelligence in Healthcare. *Int J Health Policy Manag*. 2022;12:7261. doi:10.34172/ijhpm.2022.7261

37. Rademakers FE, Biasin E, Bruining N, et al. CORE-MD clinical risk score for regulatory evaluation of artificial intelligence-based medical device software. *npj Digit Med*. 2025;8(1):90. doi:10.1038/s41746-025-01459-8

38. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342

39. Gichoya JW, Thomas K, Celi LA, et al. AI pitfalls and what not to do: mitigating bias in AI. *Br J Radiol*. 2023;96(1150):20230023. doi:10.1259/bjr.20230023

40. Filicori F, Addison P. Intellectual property and data ownership in the age of video recording in the operating room. *Surg Endosc*. 2022;36(6):3772-3774. doi:10.1007/s00464-021-08692-8

41. Ozmen BB. Navigating FDA Regulations for Development of Artificial Intelligence Technologies in Plastic Surgery. *Aesthet Surg J*. Published online April 4, 2025:sjaf051. doi:10.1093/asj/sjaf051

42. Proprio Receives FDA Clearance For the World's First AI Surgical Guidance Platform. Accessed August 13, 2025. https://www.propriovision.com/news/proprio-receives-fda-clearance-for-the-worlds-first-ai-surgical-guidance-platform

43. Technology SR. Moon Surgical Wins FDA Clearance for Maestro™ Connectivity and Predetermined Change Control Plan for AI-Powered ScoPilot®. Surgical Robotics Technology. July 3, 2025. Accessed August 13, 2025. https://www.surgicalroboticstechnology.com/news/moon-surgical-receives-fda-clearance-for-maestro-connectivity-along-with-a-predetermined-change-control-plan-for-its-ai-powered-scopilot/

44. Virtual Incision Receives FDA Authorization for the MIRA Surgical System as the First Miniaturized Robotic-Assisted Surgery Device - Virtual Incision. February 24, 2024. Accessed August 13, 2025. https://virtualincision.com/virtual-incision-receives-fda-authorization-for-the-mira-surgical-system-as-the-first-miniaturized-robotic-assisted-surgery-device/

45. CMR Surgical gets FDA de novo nod for Versius robot | MedTech Dive. Accessed August 13, 2025. https://www.medtechdive.com/news/CMR-Versius-robot-FDA-de-novo-clearance/730149/

46. Perimeter Files FDA PMA Application for its Next-Generation B-Series OCT with ImgAssist AI 2.0. Perimeter Medical Imaging. March 17, 2025. Accessed August 13, 2025. https://ir.perimetermed.com/news-events/press-releases/detail/167/perimeter-files-fda-pma-application-for-its-next-generation

47. AI Action Plan. Accessed August 13, 2025. https://www.ai.gov/action-plan

48. FDA. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions. December 3, 2024. Accessed June 3, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence

49. Health C for D and R. Artificial Intelligence and Machine Learning in Software as a Medical Device. *FDA*. Published online March 25, 2025. Accessed June 3, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

50. Stonko DP, Hicks CW. Mature AI/ML-Enabled Medical Tools Impacting Vascular Surgical Care: A scoping review of late-stage, FDA approved/cleared technologies relevant to vascular surgeons. *Semin Vasc Surg*. 2023;36(3):460-470. doi:10.1053/j.semvascsurg.2023.06.001

51. Maier-Hein L, Reinke A, Godau P, et al. Metrics reloaded: recommendations for image analysis validation. *Nat Methods*. 2024;21(2):195-212. doi:10.1038/s41592-023-02151-z

52. Qian L, Wu JY, DiMaio SP, Navab N, Kazanzides P. A Review of Augmented Reality in Robotic-Assisted Surgery. *IEEE Transactions on Medical Robotics and Bionics*. 2020;2(1):1-16. doi:10.1109/TMRB.2019.2957061

53. Byrd TF, Tignanelli CJ. Artificial intelligence in surgery—a narrative review. *Journal of Medical Artificial Intelligence*. 2024;7(0). doi:10.21037/jmai-24-111

54. Longhurst CA, Singh K, Chopra A, Atreja A, Brownstein JS. A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness. *NEJM AI*. 2024;1(8):AIp2400223. doi:10.1056/AIp2400223

55. Health C for D and R. Collaborative Communities: Addressing Health Care Challenges Together. *FDA*. Published online January 31, 2025. Accessed September 9, 2025. https://www.fda.gov/about-fda/cdrh-strategic-priorities-and-updates/collaborative-communities-addressing-health-care-challenges-together

56. Venkatesh KP, Kadakia KT, Gilbert S. Learnings from the first AI-enabled skin cancer device for primary care authorized by FDA. *npj Digit Med*. 2024;7(1):156. doi:10.1038/s41746-024-01161-1

57. American Cancer Society. Melanoma Survival Rates | Melanoma Survival Statistics. Accessed September 9, 2025. https://www.cancer.org/cancer/types/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html

58. Newlands G, Lutz C, Fieseler C. Trading on the Unknown: Scenarios for the Future Value of Data. *The Law & Ethics of Human Rights*. 2019;13(1):97-114. doi:10.1515/lehr-2019-0004

| Digital Surgery Regulatory Framework: Risk-Stratified Approach | | | | |
|---|---|---|---|---|
| **Risk Category** | **Technology Examples** | **Regulatory Pathway** | **Evidence Requirements** | **Ongoing Monitoring** |
| **LOW RISK:** Minimal potential for direct patient harm. Technologies that assist in planning, analysis, or workflow without direct procedural guidance. | • **Aidoc BriefCase**: AI reads CT scans for incidental findings relevant to surgical<br>• **Viz Vascular Suite**: AI detection of vascular pathologies for workflow prioritization<br>• **Avicenna.AI CINA Chest**: Automated detection of pulmonary embolism and aortic dissection on CT angiography | **Streamlined 510(k)**<br>• Predicated on legally marketed device<br>• Basic safety validation<br>• User interface validation<br>• Expedited review timelines | **Pre-market:**<br>• Usability testing (establish minimal acceptable performance thresholds)<br>• Basic cybersecurity assessment<br>**Clinical Evidence:**<br>• User acceptance studies | **Reporting:**<br>• Adverse event reporting<br>• User feedback collection<br>**Performance:**<br>• Basic performance metrics<br>• Annual performance summaries |
| **MODERATE RISK:** Systems providing real-time guidance requiring human oversight. Technologies that provide procedural assistance but maintain surgeon control. | • **Proprio Paradigm**: AI-guided visualization and real-time intraoperative measurements for spine surgery<br>• **Moon Surgical Maestro System:** AI-powered ScoPilot for laparoscope control during minimally invasive procedures<br>• Navigation systems with real-time guidance capabilities | **De Novo or 510(k)**<br>• Comprehensive pre-market evaluation<br>• Clinical validation studies<br>• Human factors assessment<br>• Algorithmic bias evaluation<br>• Workflow integration studies | **Pre-market:**<br>• Clinical studies (establish minimal number of procedures done)<br>• Multi-site validation<br>• Human factors validation<br>• Performance across diverse population<br>• Cybersecurity assessment<br>**Clinical Evidence:**<br>• Non-inferiority to standard of care<br>• Workflow integration validation | **Reporting:**<br>• Adverse event reporting<br>• Quarterly performance reports<br>• Real-world data collection<br>• Algorithm performance monitoring<br>**Performance:**<br>• Predefined intervention thresholds to recall from market<br>• Bias detection monitoring<br>• User competency tracking |
| **HIGH RISK:** Autonomous/semi-autonomous systems or technologies that can independently initiate irreversible actions. Novel devices without clear predicates. | • **MIRA Surgical System**: First miniaturized robotic-assisted surgery device for colectomy (De Novo)<br>• **Versius Surgical Robotic System**: Robotic-assisted gallbladder removal (De Novo)<br>• **Perimeter B-Series OCT with ImgAssist AI 2.0:** Real-time surgical margin assessment during breast cancer surgery (PMA under review) | **Premarket Approval (PMA) or De Novo**<br>• Extensive clinical trials<br>• Comprehensive safety analysis<br>• Risk mitigation protocols<br>• Mandatory training programs | **Pre-market:**<br>• Multi-phase clinical trials (n≥ X # of procedures)<br>• Safety and efficacy outcomes<br>• Failure/error mode analysis<br>• Extensive cybersecurity assessment<br>• Long-term outcomes data<br>**Clinical Evidence:**<br>• Superiority or non-inferiority<br>• Comprehensive safety profile<br>• Risk-benefit analysis | **Reporting:**<br>• Adverse event reporting (must be within X hours of critical incident) (section 522?- need to look more)<br>• Monthly performance reports<br>• Comprehensive outcome tracking<br>**Performance:**<br>• Continuous real-world evidence<br>• Mandatory registry participation<br>• Automatic shutdown triggers |
| **Key Regulatory Principles** | **Performance Standards:**<br>• Standardized metrics for each technology type<br>• Technical and clinical performance requirements<br>• Real-world validation across diverse populations<br>• Continuous monitoring with intervention thresholds | | **Safety Requirements:**<br>• Procedural risk assessment (irreversible actions)<br>• Failure/error detection and automatic fallback systems<br>• Comprehensive cybersecurity protection<br>• Mandatory user training and competency verification | |