**International Journal of Surgery Publish Ahead of Print**
**DOI: 10.1097/JS9.0000000000002309**

OPEN

# The AI-Enhanced Surgeon—Integrating Black-Box Artificial Intelligence in the Operating Room

Ronan A. Cahill, Mindy Duffourc & Sara Gerke

Ronan Cahill, M.D., Department of Surgery, Mater Misericordiae University Hospital, Dublin, Ireland

&

Centre of Precision Surgery, University College Dublin, Dublin, Ireland

Mindy Nunez Duffourc, Dr. Jur., Maastricht University, Department of Private Law & Maastricht Law and Tech Lab, Maastricht, Netherlands

Sara Gerke, Dipl.-Jur. Univ., M.A., University of Illinois Urbana-Champaign, College of Law and European Union Center, Champaign, Illinois, United States

Correspondence to:

Sara Gerke

University of Illinois Urbana-Champaign

College of Law

504 E Pennsylvania Ave

Champaign, IL 61820

gerke@illinois.edu



**Study Design**

Other - Normative/Legal Analysis

**Conflict of Interest Disclosures**

Prof. Gerke reported receiving grants from the European Union (Grant Agreement no. 101057099), the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institutes of Health Office of the Director (NIH OD) (Grant Agreement no. 1R21EB035474-01). Prof. Gerke is also a Research Fellow at the University of Copenhagen, Faculty of Law, supported by a Novo Nordisk Foundation Grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL programme - grant no. NNF23SA0087056).

Prof. Cahill receives speaker fees from Stryker Corp, Olympus and Ethicon/J&J, research funding from Intuitive Corp and Medtronic and from the Irish Government (DTIF) (in collaboration with IBM Research in Ireland) and EU Horizon (in collaboration with Arctur). He is a member of the Medical Advisory Board of Palliare and provides consultancy to Arthrex, Astellas, Diagnostic Green and Medtronic (Touch Surgery).

Mindy Nunez Duffourc is an advisory board member for the Responsible Ai UK Keystone Project on Addressing Socio-technical Limitations of LLMs for Medical and Social Computing (AdSoLve). She is also a Speinshart Healthcare Fellow funded by Hightech Agenda Bavaria.

**Funding/Support**

The authors were supported by grant 101057321 from the European Union.

**Role of the Funder/Sponsor**

The funder had no role in the preparation, review, or approval of the manuscript or decision to submit the manuscript for publication.

**Disclaimer**

The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



# The AI-Enhanced Surgeon—Integrating Black-Box Artificial Intelligence in the Operating Room

**Study Design**

Other - Normative/Legal Analysis

**Abstract**

New Artificial Intelligence (AI)/Machine Learning (ML) technology offers great potential to assist surgeons with real-time intraoperative decision-making. While currently, AI/ML-driven tools for surgeons focus primarily on technical assistance and postoperative insights, AI/ML cognitive support in surgery can add further capability. However, these AI/ML models usually conceal their underlying algorithmic reasoning process. As a result, such “black box” AI/ML models have important clinical and legal implications for patient’s safety and surgeon’s liability.

This article provides an overview of surgeons’ current practice and the potential for AI enhancement in surgical decision-making. It suggests a path toward a safe and effective integration of black-box AI/ML models into the OR. We argue that future surgeons who rely on AI for cognitive assistance do not necessarily need to fully understand, interpret, and explain the algorithmic basis of an AI’s real-time recommendation in the midst of surgery, but rather, they need to know that these tools work as promised. Assuming new black-box AI/ML models demonstrate clear benefits for surgical patients, their use will likely be incorporated into the legal standard of care and affect the liability landscape for surgeons.

**Keywords:** Artificial Intelligence, Machine Learning, Black Box, Surgery, Safety, Liability



Artificial intelligence (AI), especially machine learning (ML), in medicine has already proved useful for administrative tasks such as clerical documentation, patient scheduling, and prescribing, as well as for improving diagnostic accuracy and efficiency in radiology and pathology.[1] In surgery, AI tools are currently focused primarily on providing surgeons with technical assistance for post-surgical analysis. For example, AI technologies can use computer vision to analyze operative videos to provide valuable post-surgical insights for training, auditing, and AI development.[2] More experimental technical applications, such as automation of instrument placement or even component parts of procedures (e.g., suturing), are being benchmarked to equal human capability.[3] However, it is controversial whether using electromechanical robotic systems operated by human surgeons to provide increased dexterity has moved the needle much in improving surgical outcomes versus equivalent nonrobotic minimally invasive techniques.[4]

Perhaps the real promise of AI in surgery lies in its ability to provide real-time cognitive support for surgeons because surgery is more about 'decisions than incisions'. Fundamentally, operations involve a series of decisions, many of which are irreversible and need to be near instantaneous, with each new patient representing a new set of circumstances. AI is now capable of analyzing operative video in real-time to provide surgeons with patient-specific anatomical, physiological, and pathological insights and direction.[1] It could be that it is this type of cognitive assistance that surgeons need more than technical augmentation. However, introducing such AI surgical decision support tools into the operating room (OR) has important medical, ethical, and legal implications. In this article, we address issues related to safety and liability. We first provide an overview of surgeons' current practice and the potential for AI enhancement in surgical decision-making. We then suggest a path towards a safe and effective integration of complex and opaque AI/ML-models (so-called "black boxes") into the OR to improve patient outcomes. Lastly, in terms of liability, we argue that surgeons who rely on AI for cognitive assistance do not necessarily need to entirely understand, interpret, and explain *the algorithmic basis* of an AI's real-time recommendation amid surgery while juggling time-sensitive decisions with technical tasks that require meticulous manual dexterity, coordination, and visuospatial awareness. However, surgeons need to know, of course, that these tools work as promised, such as through the manufacturer's demonstration of successful clinical trials.

**The AI-Enhanced Surgeon**

At present, surgeons use their expertise, experience, and judgment to progress the surgery through its many sequential steps. Where technology is used, it is merely to present the surgeon with internal appearances that the surgeon alone is responsible for interpreting or to indicate the completion of cautery sealing, for example.[2] In contrast to the anesthesiologist in the same case, who has considerable patient monitoring data to support and supplement their medical decision-making, the 21st-century surgeon is essentially intellectually unaided. As a result, surgeons often operate with a high cognitive workload as they are simultaneously occupied with task performance and sequencing while reacting to unfolding intraoperative findings and tissue behaviors. While 21st-century surgeons are experts qualified to make surgical decisions using their own opaque ("black-box") reasoning, they also bear all the liability associated with adverse outcomes when those decisions are deemed to fall below the applicable standard of care.

The AI-enhanced Surgeon is a surgeon who uses black-box AI to augment and improve real-time surgical decision-making. When black-box AI/ML models like deep learning models make real-time recommendations in the OR, surgeons (and even developers) would usually not be able to understand the AI's underlying rationale or basis.[5] However, black-box AI/ML models have often been shown to be more accurate than transparent "white-box" models (e.g., decision trees) outside of surgery and thus may hold promise for improving surgical outcomes.[6] As a result, despite the lack of transparency, black box AI/ML models have great potential to catapult surgical technology beyond the current focus on using AI in the OR to improve the display and resolution of imaging systems or robotic arm movement, which have already reached their useful limit. There is no "perfect" AI tool for all surgical procedures. Instead, the future of AI-enhanced surgery lies in the development of different AI tools for different types of surgical procedures or for specific component steps within different operations.

**The Path Towards Integrating Safe and Effective Black-Box AI/ML Models into the OR**

The current regulatory approach to medical AI technologies often emphasizes the need for high levels of *interpretability* or *explainability*. This focus can limit surgeons' AI use to situations in which they can understand (or have a perception of understanding) the AI's algorithmic reasoning process.[4] This approach, however, may slow the integration of beneficial black-box AI/ML models in the OR. Instead, such models (if used) should have demonstrated some value to justify their integration into the OR, particularly in terms of safety, effectiveness, and improving patient outcomes, such as through the successful conduct of clinical trials.[6, 7] Because surgeons often do not have the benefit of avoiding or delaying crucial decisions in an invasive surgery, the black-box AI/ML model should provide the surgeon with all the medically relevant information relating to its output — what assumptions it makes and how it weights the surgical recommendation — so that surgeons can effectively exercise their expert judgment.[7] In this context, it is important to clearly define which information is medically relevant and thus must be provided to the surgeon to facilitate their expert intraoperative decision making and protect the patient's well-being. Furthermore, human factors testing and the design of these AI tools are crucial to determine whether they effectively assist surgeons' decision-making.[6] Therefore, the AI's method of presenting information to surgeons in the OR (i.e., stop-go traffic light display versus dense data ticker tape) requires careful consideration and may need to include a visual indicator of uncertainty or probability. Human factors testing should also consider cases where the AI might present an output that conflicts with traditional signals. While the technical inner workings of the black-box AI/ML model's algorithmic decision-making process may not be available to the surgeon in real-time, this may be less important as long as the model functions as promised and can be effectively integrated into surgical workflows. Like other new medical technologies, regulators, hospitals, and surgeons should focus on the results of clinical trials to ensure adequate safety and effectiveness before facilitating the transition of AI-driven technologies into surgical practice. At least currently, limitations on collecting training data can impact accuracy and generalizability of the AI tools developed.

In the absence of pure AI decision-making in surgery, surgeons and hospitals should collaborate closely with AI developers, especially when AI systems are learning and evolving. This likely requires users to provide continuous feedback to the AI developer regarding their experience and the AI's performance, as well as the developer's continuous incorporation of this feedback to improve the AI's performance. Hospitals must support and maintain this collaboration outside the usual medical device product cycle after purchase, which typically proceeds without manufacturer involvement. Of course, the robustness of the regulatory framework governing AI-driven tools of the respective country will play a crucial role towards ensuring adequate safety and effectiveness throughout their life cycle.

There is no "perfect" AI tool for all surgical procedures. Instead, the future of AI-enhanced surgery lies in the development of different AI tools for different types of surgical procedures or for specific component steps within different operations. At least currently, limitations on collecting training data can impact the accuracy and generalizability of the AI tools developed.

Education and training are also crucial components of the safe implementation of black-box AI/ML models in surgery. Medical schools and training programs should integrate AI-driven tools into their curricula, highlighting the risks and benefits of such technologies and the continued importance of human medical expertise. All surgeons using AI-driven technology should undergo extensive surgical training, either through existing training programs and protocols or through special certification training offered by the manufacturer. Surgeons should also be required to maintain and update their competencies for using AI-driven tools at regular intervals.



**Liability Aspects**

In keeping with the current medical AI technology development trend, black-box AI/ML models in development are often benchmarked against the ultimate clinical outcome, final pathology, anatomical, or physiological ground truth, but it may be better, at least initially, to ground them against a peer surgeon's interpretation or decision given a similar set of circumstance and appearances. However, it is crucial to recognize that surgery is not an exact science, and surgeons are not statisticians. Indeed, whether a surgeon is held liable for a patient's surgical injury is determined by assessing whether the surgeon's actions fall below the standard of care, which considers the generally accepted standards for surgeons in similar circumstances, regardless of the final surgical outcome. Whether a surgeon's reliance on the recommendation of a black-box AI/ML model to make a real-time surgical decision breaches the applicable standard of care is an ongoing question since that standard is *not* static but evolves.[8] Further complicating the liability issue is that the AI recommendation may affect only a single step out of the many involved in surgery (extending indeed back to preoperative factors and preparations). This would make it difficult to determine whether the black-box AI/ML model's involvement in the operation caused an adverse (or even improved) surgical outcome for the patient. In the future, however, rather than assume all liability for decisions and actions in surgery, a surgeon whose real-time decision-making is enhanced by cognitive assistance from a properly developed black-box AI/ML model would ideally share liability risks with the AI developer and hospital. This would allow a fair allocation of liability that can take into account potentially substandard conduct of both the manufacturer and the healthcare organization under existing medical liability principles. Figure 1 below describes black-box AI/ML models and summarizes the main recommendations for best integrating such models in the OR.

**Figure 1: Integrating Black-Box AI/ML in the OR**

[Insert Figure 1]

**References**

1. American Medical Association, Future of Health: The Emerging Landscape of Augmented Intelligence in Health Care, https://www.ama-assn.org/system/files/future-health-augmented-intelligence-health-care.pdf (2024).

2. Mithany RH, Aslam S, Abdallah S, Abdelmaseeh M, Gerges F, Mohamed MS, Manasseh M, Wanees A, Shahid MH, Khalil MS, Daniel N. Advancements and Challenges in the Application of Artificial Intelligence in Surgical Arena: A Literature Review. Cureus. 2023;15(10):e47924. https://doi.org/10.7759/cureus.47924.

3. Amin A, Cardoso SA, Suyambu J, Abdus Saboor H, Cardoso RP, Husnain A, Isaac NV, Backing H, Mehmood D, Mehmood M, Maslamani ANJ. Future of Artificial Intelligence in Surgery: A Narrative Review. Cureus. 2024;16(1):e51631. https://doi.org/10.7759/cureus.51631.

4. Kawka M, Fong Y, Gall TMH. Laparoscopic versus robotic abdominal and pelvic surgery: a systematic review of randomised controlled trials. Surg Endosc. 2023;37(9):6672-6681. https://doi.org/10.1007/s00464-023-10275-8.

5. Gerke S. “Nutrition Facts Labels” for Artificial Intelligence/Machine Learning-Based Medical Devices—The Urgent Need for Labeling Standards, 91 Geo. Wash. L. Rev. 79 (2023).
6. Gerke S. Health AI for good rather than evil? The need for a new regulatory framework for AI-based medical devices. Yale J Health Policy Law Ethics. 2021;20(2):433–513.
7. Bellini V, Russo M, Domenichetti T, Panizzi M, Allai S, Bignami EG. Artificial Intelligence in Operating Room Management. J Med Syst. 2024; 48(1):19. doi: 10.1007/s10916-024-02038-2.
8. Duffourc M, Gerke S. Decoding U.S. Tort Liability in Healthcare’s Black-Box AI Era: Lessons from the European Union. Stan. Tech. L. Rev. 2024;27:1-70.

Figure 1