Operational framework and training standard requirements for AI-empowered robotic surgery

O'Sullivan, S.; Leonard, S.; Holzinger, A.; Allen, C.; Battaglia, F.; Nevejans, N.; ... ; Gallagher, A.G.

DOI: 10.1002/rcs.2020

REVIEW ARTICLE

The International Journal of Medical Robotics and Computer Assisted Surgery

WILEY

Operational framework and training standard requirements for AI-empowered robotic surgery

Shane O'Sullivan[1] | Simon Leonard[2] | Andreas Holzinger[3] | Colin Allen[4] | Fiorella Battaglia[5] | Nathalie Nevejans[6] | Fijs W. B. van Leeuwen[7] | Mohammed Imran Sajid[8] | Michael Friebe[9] | Hutan Ashrafian[10] | Helmut Heinsen[1,11] | Dominic Wichmann[12] | Margaret Hartnett[13] | Anthony G. Gallagher[14,15]

[1]Department of Pathology, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

[2]Department of Computer Science, Johns Hopkins University, Baltimore, Maryland

[3]Holzinger Group, HCI-KDD, Institute for Medical Informatics/Statistics, Medical University of Graz, Graz, Austria

[4]Department of History & Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania

[5]Faculty of Philosophy, Philosophy of Science and the Study of Religion, Ludwig-Maximilians-Universität München, München, Germany

[6]Research Center in Law, Ethics and Procedures, Faculty of Law of Douai, University of Artois, Arras, France

[7]Interventional Molecular Imaging Laboratory - Radiology department, Leiden University Medical Center, Leiden, the Netherlands

[8]Department of Upper GI Surgery, Wirral University Teaching Hospital, Birkenhead, UK

[9]Institute of Medical Engineering, Otto-von-Guericke-University, Magdeburg, Germany

[10]Department of Surgery & Cancer, Institute of Global Health Innovation Imperial College London, London, UK

[11]Morphological Brain Research Unit, University of Würzburg, Würzburg, Germany

[12]Department of Intensive Care, University Hospital Hamburg Eppendorf, Hamburg, Germany

[13]GB Group plc, Chester, UK

[14]Faculty of Life and Health Sciences, Ulster University, Londonderry, UK

[15]ORSI Academy, Melle, Belgium

Correspondence
Shane O'Sullivan, Department of Pathology, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Av. Enéas Carvalho de Aguiar, 155 - São Paulo/SP - Cep 05403-000, Brazil.
Email: doctorshaneosullivan@gmail.com

Abstract

Background: For autonomous robot-delivered surgeries to ever become a feasible option, we recommend the combination of human-centered artificial intelligence (AI) and transparent machine learning (ML), with integrated Gross anatomy models. This can be supplemented with medical imaging data of cadavers for performance evaluation.

Methods: We reviewed technological advances and state-of-the-art documented developments. We undertook a literature search on surgical robotics and skills, tracing agent studies, relevant frameworks, and standards for AI. This embraced transparency aspects of AI.

Conclusion: We recommend "a procedure/skill template" for teaching AI that can be used by a surgeon. Similar existing methodologies show that when such a metric-based approach is used for training surgeons, cardiologists, and anesthetists, it results in a >40% error reduction in objectively assessed intraoperative procedures. The integration of Explainable AI and ML, and novel tissue characterization sensorics to tele-operated robotic-assisted procedures with medical imaged cadavers, provides robotic guidance and refines tissue classifications at a molecular level.

KEYWORDS

surgical skills, dexterity, autonomous robotic surgery, supervised autonomy, explainable artificial intelligence xai, surgical navigation

1 | INTRODUCTION

Our thesis statement is 3-fold. Firstly, morphological and molecular medical images of cadavers could potentially be used to evaluate skills training in autonomous robotic surgery. Secondly, artificial intelligence (AI) can enhance robotic navigation, tissue diagnosis, treatment, and clinical management during surgery. Thirdly, combining surgeon-in-the-loop and computer intervention into the decision-making process complements their respective strengths. However, we take into consideration that the detailed differences between a living person and a cadaver are not to be underestimated. Moreover, a good and safe surgery is very much about the detail of performance.

1.1 | Specific aim and methodology of this paper

The specific aim of this paper is to provide an initial step toward an operational framework and training standard requirements for AI-empowered robotic surgery. Only through the development of standards that ensure ethics and safety will autonomous robotic surgery actually make it mainstream.

This review was based on a comprehensive search of relevant published scientific literature found on the PubMed and DBLP databases. In the case of Pubmed, we used MeSH tools to search the terms: (autonomous robotic surgery) OR (surgical robot) OR (supervised autonomy) OR (training model) OR (surgical navigation) OR (dexterity). We manually searched for studies in the list of references in the review articles, particularly studies that used tracing agents/contrasts/fiducials. We additionally searched PubMed and DBLP databases to include studies related to augmented reality visualization and transparency. To support our rationale for transparency, we included studies that investigated explainable AI (XAI), machine learning (ML) black box solutions, and algorithmic transparency/opacity. After selecting the publications, we used the information to draw discussions under two main sections entitled: "Challenges and knowledge gaps"; and "Hypotheses and recommendations."

1.2 | Gross anatomy

In academia, gross anatomy training courses are an untapped, highly fertile, source of knowledge, and direction for AI and autonomous robotic skills training. We are putting the spotlight on this untapped source because it can be a starting point to open new avenues to develop and advance autonomous robotic performance with explainable outcomes. It is indisputable that gross anatomy training plays a major role in enhancing medical research and education. It is empirically proven that simulations, if designed and deployed adequately, can significantly enhance learning in medical education.[1-5] Today, cadavers are still used to verify surgical techniques prior to surgery on living patients.[6] Although such practices are not worldwide, in the United States, appendectomies are still practiced on human cadavers and not in some computer technology simulations.[7]

1.3 | Autonomous surgical robots

Developments in autonomous robotic surgery are appealing to United States military planners[8] and the Food and Drug Administration (FDA).[9] In terms of public health care or general hospitals, autonomous robotic ENT (ear, nose, and throat) procedures are particularly desirable and beneficial for patients with high-risk pathogens[10] (eg, coronaviruses/COVID-19). In otorhinolaryngology (ORL) therapy procedures, micro-robots could also be used to operate in complex scenarios involving restricted spaces within the face.[11,12] For instance, the nasal cavity, eye orbit, and mouth areas have multiple nerves and structures that can easily be damaged. Autonomous microrobotic surgery could prevent surgical injuries within these very confined anatomical spaces (this also applies to ophthalmology). Typical injuries in confined spaces include spinal nerves in spinal surgery, great vessel (aortic, vena cava, and pulmonary artery) injury in cardiothoracic and vascular surgery, as well as pelvic plexus, hypogastric nerve, and pudendal nerve injury in prostatic/pelvic approaches. This also applies to ophthalmology. Superpositioned in a very small space, microrobots can facilitate the novel removal of tumors and even cope with complicated retinal pathology.[13]

In gynecology, autonomous robotic surgery could also enhance operating in complex surgical planes.[14,15] Gynecological surgery can be difficult for surgical calibration and access. At the pelvic rim, stabilizing surgical maneuvers with superimposed computer-generated images (generated from radiological and information-laden diagnostic sources) fulfilling enhanced augmented-reality on robotic-derived images of the gastrointestinal tract could be a solution for some complicated operations. Moreover, in orthopedics, autonomous platforms could allow enhanced reconstruction of broken limbs and joints.[10] This could achieve ideal optimization in a biomechanical approach. Space exploration or disaster management also illustrate realistic fields for autonomous robotic surgery. However, ethical issues are raised by increasing the levels of autonomy for surgical robots.[10,16]

1.4 | Using cadavers to address the problem of robot training limitations

Autonomous robots for surgery (fifth generation as described in Figure 1) cannot be expected to match or exceed human skills if access to cadavers is restricted solely to training humans. Moreover, human cadavers present a very different proposition than work on animals or training with phantoms (ie, plastic human models). One cannot program a robot based on general assumptions or properties, such as that which might be built into a simulation. Rather, robots need a broad, experientially acquired knowledge of the cadaver morphology. By combining experience of the operation with the precision of the robot and high-level molecular characterization of the tissue, one can achieve, not only the objective, but also an explainable outcome, and as a relevant practically important result, a replicable one. Robotic training with cadavers can prompt radical innovation capabilities of AI researchers that will expand the known art and enable others to pursue commercial improvisation, realization, and explainable AI itself. As with gross anatomy for training humans, cadavers can also provide robots with the opportunity to have a full-contact machine learning (ML) environment.

Furthermore, through cadaver studies that include medical imaging and other relevant data acquisition sensorics, we contend that one can assess how a human or robot may differ in skills to demonstrate consistency and autonomy in surgery (this may require tracking of the robot). Work on autonomous surgical robotics has been cited by authors who are solely interested in comparing the skills of "human surgeons" against "robot surgeons" (eg, the 2016 surgical results of the Smart Tissue Anastomosis Robot [STAR][18]co-developed by Leonard et al[19] - STAR robot is shown in Figure 2). A recent survey of other similar systems is presented in Opfermann et al.[20]

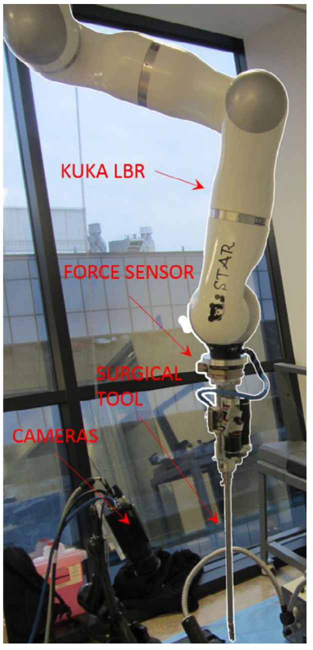

FIGURE 2 The smart tissue anastomosis robot (STAR)

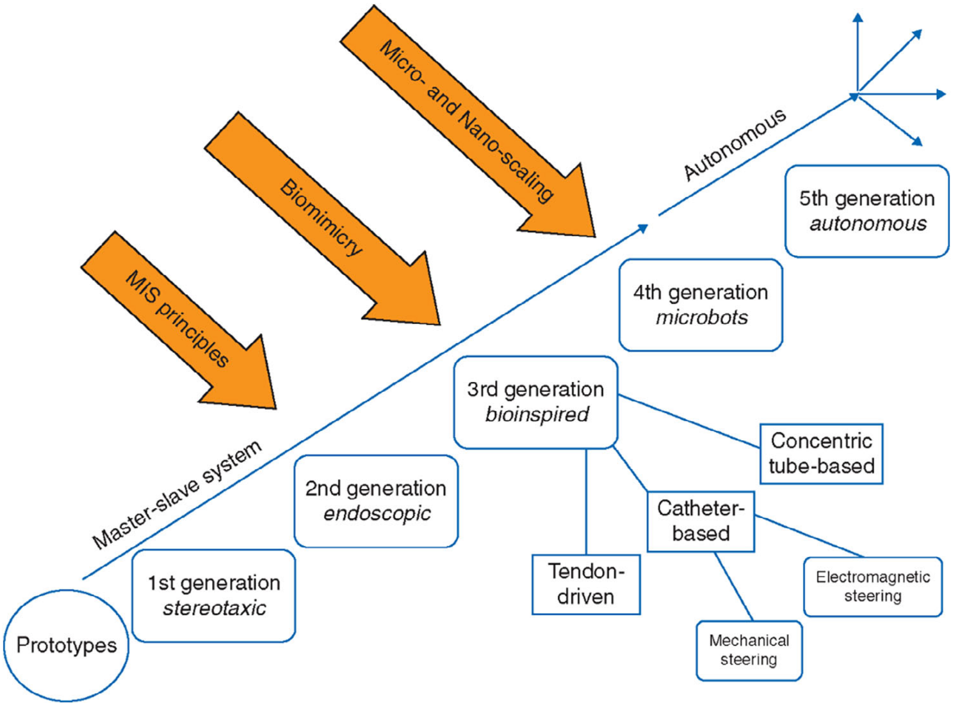

FIGURE 1 The learning curve of surgical robotics. *Source*: Ashrafian et al[17]

2 | CHALLENGES AND KNOWLEDGE GAPS

In cadaver studies, autonomous robots can develop the low-level skills and dexterity to execute a specific task in surgery. So far in AI, most work covers milling, cutting, and suturing skills. One of the motives in addition to advancing autonomous robotic procedures for surgery is to compare the "robot's performance" against that of the "human's performance" (eg, needle positioning, tissues analysis, suture quality, speed, etc). From a research standpoint, it is worth evaluating the variability of the performance results to determine factors such as how often humans fail, or how costly human procedures are, among other factors. Therefore, these are important knowledge gaps for further studies. For these cadaver studies, one may encounter restrictive access to costly high-end imaging devices such as computed tomography (CT). However, it may be more practical to use interventional tools (eg, conebeam CT, ultra sound, and fluorescence) coupled to the robot.

It is worth mentioning that both humans and the robotic system have problems accurately guiding devices to an identified area for collecting tissue samples for subsequent pathological examination, for example. While diagnostic imaging is generally able to show a pathological or otherwise interesting area, the device itself—in our case robotic probe—causes imaging artifacts when inserted into the human cadaver. These artifacts can lead to significant distance errors that subsequently lead to inaccurate determination on whether the target site has been reached or not. Biopsy samples from the wrong target site could lead to false negative results. We, therefore, propose that one can add audio sensors to the robotic systems that measure the signal between a tool-tissue interaction, and other similar existing approaches should be taken into consideration.[21] Advanced signal processing could extract features that can be used to improve device guiding accuracy and may even be able to classify tissues.[22,23] The obtained cadaver audio profiles of different tissues could be used as base information to create a virtual tissue histology database. In

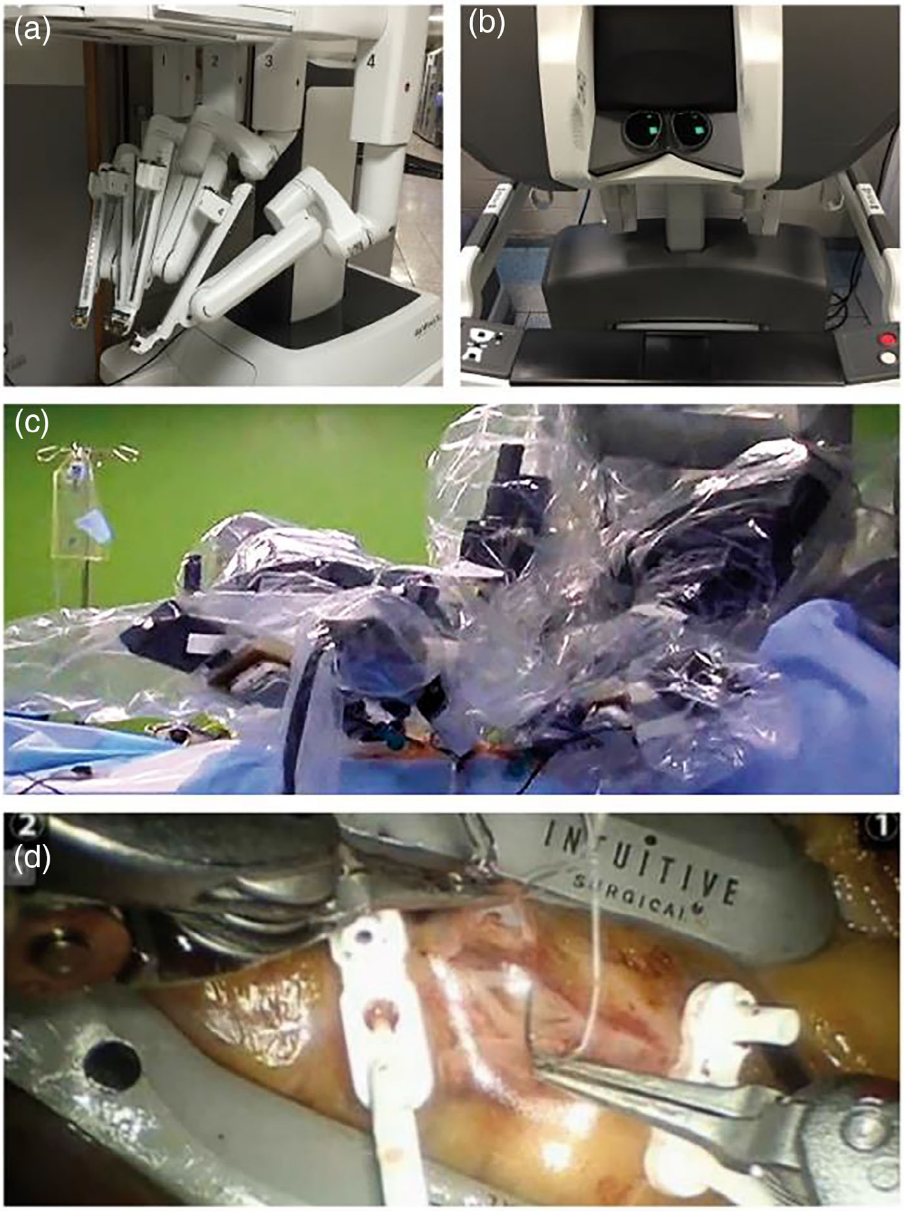

FIGURE 3 The da Vinci Xi Surgical System A, robot arms, B, console, C, operation setting, and D, coronary anastomosis surgical procedure. *Source*: Ashrafian et al[17]

remote robotic surgery, audio sensors have been added to the Da Vinci (see Figure 3) to simulate and "display" haptic and palpation information.[24]

2.1 | Cadaver training with tracers/agents and perfusion devices

Autonomous robots, trained to perform surgery without depending on contrast agents, would have a variety of practical positive uses and advantages. Many fluorescent markers or tracers require FDA approval and biocompatibility. This is very much restrictive for robotic surgeries on living humans (it may even require ethical approval to use such agents during training on live animals). One could focus on navigation of the robot by using 3D imaging and molecular tissue characterization using multiplexing fluorescence immunohistochemistry approaches (the analysis of such multiplexing data could benefit from AI). This way one can develop generic technologies that could later also be applied during surgery or pathology, respectively.[25] Nevertheless, when using nonapproved potent agents on human cadavers (eg, in angiography), toxicity is not an issue for surgical training and robots can dissect with relatively higher accuracy. Yet, one can question how to deliver a tracer to a hidden organ in a cadaver without intact blood supply? Artificial circulation (eg, with a perfusion device) is a possible solution for this challenge, but this may only be realized with lipophilic tracers, and in this case, fluorescence can be used for the robot's tissue identification task. (See the impressive work of van den Berg et al with the da Vinci robot that used a fluorescent tracer on humans.[26]) Decker et al used the same fluorescent dye, but as fiducial markers instead of a tracer for autonomous procedures on cadavers.[27]

Nonetheless, this postmortem setting is still ideal for testing new tracing agents and developing a new generation of AI-enhanced robotic surgeons. However, one could question the proficiency of robots when they are confronted with individual variability of organs and systems. Therefore, recorded tracking of the robot relative to pre-intervention images would be appropriate. The topography of the arterial system can even be a challenge for many anatomists. There is a substantial anatomical variation between individuals, and this is often related to the variety of arterial distribution. Therefore, imaging after perfusion with, for example, radiopaque contrast, could help identify critical deviations between the anatomy of the cadaver and that in textbooks. Same will be true for rigid structures such as bones. In combination, the two could help to realize a deformable registration. Like an experienced anatomist (or surgeon), a robot can remain calm in a surge. However, for autonomous robotic surgery, one of the challenges will be to integrate bowel movements or arterial pulsation.

2.2 | Skills in surgery

Nowadays, authors reviewing surgical robotics are putting much emphasis based on comparing surgical skills of "humans" vs "robots." One could argue that the da Vinci robot (the leading tele-operated-robotic-assisted surgery system[17]) is mainly assessed in urology and that its results are only measured at a macroscale (eg, reduced complications). This implies that microscale assessments are excluded, such as suture quality. This usability is perhaps a major reason why surgeons want to use a robot. Without wristed tools, laparoscopic suturing can be quite complex. But this depends on the surgeon because many skilled surgeons do not need a dextrous wrist. Either way, using a dextrous wrist vs straight tools should not preclude the assessment of small-scale performances.

When examining how surgeons assess a suture, one must aim to find specific characteristics.[28] One could attempt to gather elements from previous studies to define what makes one suture more desirable than another suture. This could suggest that one should analyze current robotic procedures. Here, tortuosity is considered to be a key skill-related feature. However, dextrous/tortuosity features of hardware should not dictate how a (sub)task can be objectively assessed. Today, it seems challenging to find surgeons that can objectively determine what makes a good suture and knot tying other than a subjective assessment. It is important to have an objective standard of what makes a "good" suture or how the quality of a suture can be objectively measured.[28,29] Metrics for hand movements may not provide tracking quality, since hand track is a measure of process, not quality. For example, two surgeons can have different path lengths for the same surgical procedure. One surgeon could have had a more difficult case or the other surgeon simply did not perform parts of the procedure. Besides focusing on sutures, we also see relevance in studying biopsy accuracy, needle placement during ablations, or resection margins.

2.3 | Surgery as a science rather than an art form

One may determine that a suture is good when it makes what it is designed for, the closure of tissue. Surgeons with lower rates of leakages are considered better surgeons. Yet, these may still be considered fairly subjective assessments; however, there is now a considerable body of robust evidence demonstrating across disciplines and procedures that interventions can be validly characterized in detail. In addition, most tissues will leak right away. In bowel anastomosis, surgeons will not witness blood or leakage until digestion resumes. One could doubt that leakage rates of sutures are properly assessed during surgery other than visually. The difficulty in finding straightforward answers makes this an art form rather than a science form. If a suture appears acceptable and feels good, then it is favorable. Some surgeons may evaluate the skills of a surgeon based on their number of useless movements and their magnitudes.

2.4 | Defining performance standards for autonomous robotic surgery

It is necessary to evaluate the variability of results for "autonomous robot" vs "human." The lack of standardization of these results is a major knowledge gap. Cadaver training research potentially fills this vacuum by proposing metrics and methods to standardize the assessment of accuracy, consistency, and efficiency of surgical robots. One can address this through automation of robotic sampling. Sending both robots and engineers into this fray automatically brings forth the desire for "standardization." Although there are few "standards" to assess how well a system performs, O'Sullivan et al highlight that there are several datasets that are used for benchmarking.[10] In robotics, there is an engineering tradition of establishing a clear standard to define and categorize performance (accuracy, repeatability, etc.), but also standards for procedures to measure them. Therefore, purchasing a robot that conforms to standards (eg, ISO XYZ), we know exactly what it can or cannot do (ie, execute XYZ with a standardized measure of accuracy and repeatability).

There is a broad range of skill set among different human surgeons. Some human surgeons are more dexterous than others at performing certain tasks, whereas robot automation leads to movements with less variability, resulting in more consistent and predictable results. Automation standardizes results to a significantly higher degree than humans—trial after trial produces similar positive results. In robotics, researchers can receive criticism when making attempts to define an ad hoc measure of accuracy for it. For example, there is no standard definition of "accuracy" for navigation systems commonly used in neurosurgeries or how accuracy must be measured. A "standard" should be defined regardless of some challenges such as the variability of the human anatomy.[30] However, one approach could be to focus on a standard that could improve data recording and coupling to machine learning. Therefore, a standard could be the forensic accuracy, because just as in surgery, pathology can also provide the gold standard.

2.5 | Criticism of robot navigation

There are several robots autonomously navigating buildings albeit they might not be as proficient as humans and require floor plans or explorations. Criticism goes beyond this type of navigation. Take the case of the robot navigations through the nuclear plant following the Fukushima Daiichi nuclear disaster in 2011. There is a common misconception that these robots were navigating autonomously, but instead they were actually tele-operated. However, this recent example may not be considered a normal navigation scenario. These robots experienced malfunctions because of the strong radiation, poor communication environment, or were limited due to their incapability to navigate stairs.[31] Nevertheless, there are plenty of challenging environments where robots navigate successfully. For example, Minerva, a mobile robot, offered guided tours of the Smithsonian Museum of American History in Washington, DC, for 14 days in 1998. During that period, the robot offered 620 tours and traversed a distance greater than 44 km through the museum's crowd.[32]

2.6 | Human thinking, gender, racial, and social biases

We question whether one can train algorithms to expose hidden biases in such systems. "Black-boxed" or opaque algorithmic processes can perpetuate and reinforce, morally and epistemically, harmful biases. For example, Vallor and Beckey[33] highlight that it is very common that human thinking biases (eg, racial, gender, or socioeconomic) become embedded in human-generated datasets, which are used to train or educate robot systems. These data define the "world" that an AI agent "knows." Yet the effect of human biased data on robot outputs is easily obscured by many factors, making those biases more harmful and resistant to eradication.[33] A robot might, for example, be subtly slower or a bit less precise given situations that were rare in its training set, and the effects of these small differences may be hard to detect on a case-by-case basis, while still contributing to biased outcomes when analyzed at the population/subpopulation level. Moreover, too many samples with pathologies may also lead to overdiagnosis by AI systems when challenged with "real world" data.

Even researchers who understand the mechanisms by which human bias can infect computer intelligence are frequently surprised to realize the degree of such bias in computer outputs—even from inputs considered to be relatively unbiased.[34] Bolukbasi et al[34] also highlight important aspects of gender biases.

Learning data can also be based on religious, racial, or social biases introduced by designers or trainer surgeons. These risks are mentioned in the predictive justice sector in the United States. Biases are not always intentional. The algorithm itself may not be considered racist/prejudiced. The problem is that trainers and software testers may not have taken care to use ethnically diverse data.[35] These risks may also exist in surgical robotics.

2.7 | AI robots potentially possess or lack overconfidence (surgical ego)

For the expert-in-the-loop, gender factors are important to consider because survey evidence shows that male doctors engage in more disruptive behavior and there appears to be better outcomes for the patients of female surgeons than those of male surgeons.[36,37] We question whether the machine tends to perpetuate this present condition or whether robots can be designed and configured in ways that promote gender equality. By investigating this for AI, it is necessary to question whether AI surgical robots would be more preferable than an occasional overconfident surgeon.

AI robots potentially lack overconfidence because "surgical ego" depends on who educates and trains the robot. The "surgical ego" may lead to irrational reasoning and recommendations in favor of performing complicated, risky, and unnecessary laparoscopic surgeries, simply to prove that they are possible.[36,37] In other words, the surgical egos of surgeons used to model actions for the machine to learn from, may lead to the robot acquiring risky behaviors. There may also be some difficulties in developing ways to change this from happening, since this may require a deep "cultural change." Therefore, this is not exactly the ideal or appropriate learning environment for an AI robot that is learning from a surgeon (eg, movements, tissue interpretation, and more) and coding their skills. One could expect that surgical robots possess patience and avoid rushing operations or taking inappropriate shortcuts.

It is also worth asking how to actually determine if robots are learning from suitable surgeons or from egotistical/stressed surgeons? Stress has been cited as an excuse for ego problems and "surgical ego" is dangerous because it drives surgeons to take short cuts or hide poor organization by performing fast incomplete surgeries.[36] Therefore, it is important to know who exactly is training and educating these robots.

2.8 | Surgery and robot learning on a cadaver, explainable results, and norms

The use of surgical robots can lead to specific risks for patient safety,[38] especially the risk of perforation, injury, or tissue burn.[39] It could be physically difficult to be aware of these disadvantages by training on a cadaver. The risk would be then that the robot having learned inappropriate information about a cadaver still reproduces the same errors on a patient. It is essential to ensure that the development of autonomous surgical robot learning does not increase existing risks in surgical robotics. However, the scientific literature shows that the training of the surgeon is decisive in reducing risks on robotic-assisted surgery[40]: the better the surgeon trains, the more complications are reduced. The same would be true for robots.[41-45] Another limitation relating to machine learning techniques is also to think about the risks for the patient. To avoid putting the life of the patient in danger, the robot is required to be able to inform the surgeon operator in the case of a delicate/unknown situation or seek the surgeon's approval for a course of action.

We see in a general way: (a) the algorithm needs to be explainable also for the doctor who must be able to check at any time what the robot has learned; (b) it is important to set a protocol for the surgeons who train the robots, to prevent machines from learning inappropriate actions. This should involve thinking about their own practices in order to rationalize the right actions and to fix them in a protocol[41,46]; (c) the surgeon operator has to be informed when the robot has not been trained on a specific case; and (d) future standards will need to consider these limitations and anticipate them.

2.9 | Confronting several problems with explainability

We do not suggest that automation should, or will, replace conventional or robotic-assisted surgical tasks on a short notice. We simply postulate that using cadavers for developing explainable AI robots fosters and complements robotic-assisted procedures in general. It can be considered that it would only be the "experience" or a "metric driven experience," that is, procedure events which are operationally defined. Today, designers and developers of AI/machine learning seek new ways to make machine reasoning more transparent, and to avoid biases in the input data and training sets.

As a matter of fact, the most successful current machine learning methods are considered as so-called "black-box"-approaches.[47] While it is not completely correct to refer to such methods as "black-box," because we know the mathematical principles behind such approaches, the inside-complexity makes it difficult to identify the important explanatory factors underlying outcomes in a specific context of a problem. In practice, this means that the results of specific input to output transformations cannot easily be traced back, hence such approaches are defying easy causal analysis.

With respect to explaining the basis of decisions, O'Sullivan et al[10] highlight the current situation. Compared with traditional logic-based or symbolic AI approaches, successful current methods such as "Deep Learning" (many layers of neural networks) are statistical learning approaches, which are highly nonlinear and require enormous amounts of training data.[48] The central problem here is that this is conducive to an epistemically opaque relationship between input and output. This opacity can result in a loss of understanding.[49]

If humans cannot reliably query an AI robot about the basis and reasons behind its decision-making, then how can humans reliably assess the decision's validity?[50] Today, the problem of systematically evaluating explainable AI confronts several problems: (a) it assumes a "gold standard" for explanations that may not exist given the current state of the art; (b) it requires the input of experts, who may have biases leading them to question and reject explanations that are outside their experience; and (c) it needs to be calibrated to the background knowledge, social context, and ethical expectations of users, be they physicians, other medical workers, patients, or relatives of patients but also to specific needs of the Courts in the event of a lawsuit against the hospital, surgeon, or robot manufacturer. Usually, a systematic approach to these challenges would require: (a) the adaptation of methods of model testing, such as withheld data testing, from their usual context of prediction to that of explanation; and (b) a clear statement of the target audiences for explainable AI, and the development of different strategies that acknowledge the various goals and beliefs of those different audiences.

Explainable AI is a new area of research, but it is gaining momentum. Explainable AI is used in the regulatory and compliance area of healthcare. However, interviews with experts in these fields (who in the main are non-technical) revealed, in several instances, a mistaken conflation of the explanation of decision-making with the explanation of algorithm training. The challenges of explaining algorithmic based decision-making are not unique to AI. Explaining multivariate statistically based models and inferential models to non-experts is just as fraught.

The challenge of explaining algorithm training is disturbing in many ways; it reflects the reality that open-source AI tools are already readily available for anyone to use regardless of any skill or professional training. The procedures and safe-guards that are commonplace to AI researchers (eg, to address training set bias) are not typically used by amateur enthusiasts. Thus, the credibility of a trained AI tool may be suspected. But this is not a failure of the algorithm itself, rather it is a failure of the humans who trained it. While it may seem like an issue of semantics, it is not. The failure of humans to use professional standards and procedures when training an AI can be readily detected by another professional AI expert with some directed questioning. However, more fundamental failures in an algorithm are much more difficult to detect, because the algorithm cannot be interrogated in the same way as a human.

2.10 | Overriding AI

To apply the questioning of Vallor and Beckey[33] to a surgical case, one can question: if an AI surgical robot has consistently demonstrated higher competence than humans in specific tasks, then on what grounds would an egoistic/overconfident/narcissistic surgeon be given the power to challenge or override its decisions? This issue is for explainable AI to prevent and minimize any conflicts that may occur. AI can provide diagnosis and recommendations.[10] However, conflict may occur between the surgeon in charge and other staff, who may be in favor of the AI recommendations. Surgeons in charge may fail to accept that their own diagnosis or therapeutic plan is erroneous or ill-conceived, or decline the AI diagnosis or therapeutic alternative.[36]

2.11 | Favoring AI

On the other hand, the opposite may occur. Boscarato argues that informed consent could be susceptible to influence by hospital staff. Buying and installing surgical robot systems are very expensive processes.[51] Hospital staff may try to persuade patients to accept the robotic procedure in order to pay off the costs of buying and installing the system. Nevejans emphasizes that consent must, therefore, relate not only to the surgical procedure itself, but especially to the use of the robot.[52] Therefore, regulating informed consent is very important for providing patients with procedure details, explanations for robot use, potential risks, side effects, and the advantages against using conventional surgery.[51] A European search report also requires the doctor to indicate the sources that would make his/her claims legitimate, so that the patient can assess their reliability and decide to use conventional surgery.[53] The use of surgical robotics requires the patient's consent.[52] Moreover, even in cases whereby a robotic surgery is more preferable than a conventional surgery, the patient may have personal reasons to refuse it, such as a lack of confidence in the robot or in the algorithm, or simply being unable to pay the extra cost. All of these factors need to be respected by the surgeon and hospital staff.[52]

2.12 | Delegate the medical decision to the AI

Another phenomenon could occur regarding the increasing decision-making role of the autonomous surgical robot in dealing with humans. This effect can already be seen in medicine with the development of decision support algorithms. Nevejans shows that they can impact the care relationship and push for a form of delegation of the medical decision to the AI, because the doctor risks: (a) to withdraw from the decision-making process in front of the algorithm whose response is supposed to be better; (b) to be unable to explain the result of the algorithm to his patient, to inform him, and to obtain full consent; and (c) to feel relieved of liability for an algorithm which makes it seem like deciding in his place.[52] In surgical robotics, to avoid this delegation of the decision to the AI, it would be necessary to put in place a protocol to make sure that the human retains autonomy relating to clinical matters. For this purpose, the robot should detail the surgical process chosen for the patient, and allow the surgeon the opportunity to preview the various stages of the operation and to validate the operative process of the robot if the doctor believes it is adequate.

3 | HYPOTHESES AND RECOMMENDATIONS

Prior to training an autonomous robot on human cadavers, we recommend that one should test the robot on organs and tissues of euthanized large animals (eg, pigs). Human cases can begin later under the supervision of medical experts (eg, surgeons). In what follows, we provide hypotheses and recommendations to guide others on their cadaver work and deployment of standards for training autonomous surgical robots. We submit that morphological imaged cadavers and tracking technologies should be used to train robots to navigate and react in real life to relevant robotic cadaver scenarios. This approach can facilitate filling some of the knowledge gaps in the previously discussed: (ie, “Challenges and knowledge gaps” section of this paper).

3.1 | Mapping recorded training/pre-interventional information to the reality

Tele-operated robots that assist surgical procedures, in synergy with learning-algorithms, enable the robot to learn from the surgeon (eg, robot arm positioning, tissue interpretation, and more) and code these skills. To devise the optimal learning-algorithms, we recommend a proposition point: the robot's recorder data (eg, telemetry/command, camera footage, etc.) should be processed and analyzed during training and performance evaluation. This data processing and analysis could lead to autonomous surgical procedures, augmented reality visualization (eg, with ghost overlay), and alert systems for procedure guidance. For example, on-screenreal-time faded guidance markers and/or split screen footage of similar previous procedures. Such an approach could provide alerts regarding errors or potential dangers.

Just as relevant to standardization, it can also provide skills' evaluation of “novice vs expert” and even “human vs robot.” Realistically, at some research/training centers, the robot could learn and know the plans of many specific surgical procedures. However, at the beginning of the operation, the anatomy can slightly change and organs may move. With one robot arm holding a needle guidance or imaging device (eg, ultrasound transducer), the robot could be able to position itself close enough to obtain a real-time image of the anatomy and update its recorded-training (and pre-interventional) information of the planned operation.[54,55] This AI robot can detect that its recorded-training (and initial) information has changed after incisions or dissection have been made. At certain points, it can then detect that it needs to change its procedure. The robot could also be instructed about where it is not allowed to dissect by introducing virtual walls that oppose resistance when its instrumentation is too close to areas of the anatomy.

For planned conventional surgeries, human surgeons sometimes view a screen with a 3D image of the patient. However, this image may have been obtained days or even weeks prior to the operation. Human surgeons use their hands or surgical senses (eg, even smell) and their own knowledge to perform the operation. We contend that for AI robots, tracking combined with their camera view and state-of-the-art tissue sampling methodologies can allow them to see the patient and precisely map the pre-interventional information to the reality.

3.2 | Imaged cadavers with tracers/agents for skills training and standardization

Automating postmortem studies has major significance and opens new opportunities for testing automation using new tracing agents. It provides an ideal test bench for testing novel near-infrared or radioactive tracers,[26,27] that are fundamental to automate surgery procedures and, compared with testing in live patients, entails a reduction in the level of clearance required by regulatory boards. Most agents need viable tissue to work with. At times, postmortem can be an issue; however, tests with immunohistochemical stainings could be performed.

For the foreseeable future, autonomous robots cannot be expected to perform autonomous procedures effectively without using a combination of morphological and molecular imaging for guidance. It is certainly worth considering this approach for developing and testing AI-driven procedures in frequently repeated cycles of optimization. One can create multiple different scenarios. Scenarios such as needle placement and sample analysis could have clear benefits and could also be considered less technically complex in comparison to other scenarios such as dissecting open, surgical removal, and suturing closed; various endoscopic navigations; or closed reduction surgery.

One area needed to be advanced is autonomous navigation; AI can address this challenge by planning a robot motion to navigate to the source of a problem that may have been overlooked by a human. AI is very promising in radiology, so AI can help identify the target, and initially this can be done together with a human operator.[56] This is a real prerequisite that one will find compelling and lead to a great result. Novel integrative and interactive machine learning approaches are complemented and extended by an expert surgeon integrated into the robot learning procedure.

One intention could be to perform, for example, CT imaging before and after the cadaver training for "automation standardization" or for making comparisons of the performance attributes of "robots" against "humans." Using automation, one can provide equal or superior results even when compared with state-of-the-artrobotic-assisted surgery devices. It is worth proposing to train autonomous robots on how to prepare, position, and react to organ/bowel movements in an open surgery scenario—which is a topic for machine learning[57] and can be facilitated using image navigation and tool tracking. Nevertheless, it is still just as important to focus on cadaver work for needle navigation, and in this case, the subject may remain closed. A biomimetic learning approach could include a robot imitating human hand movements, however, on its own, this approach could be considered dangerous. Yet, it could be more efficient to develop skills for actuated tools. This learning approach can make certain skills obsolete (eg, needle transfer).

3.3 | Cultural skills, legal rules, and avoiding biases

Gross anatomy training is usually taught to clinical students who, for the first time in their life, will be confronted with cadavers. For better training, university staff may communicate cultural skills instead of reproducing facts. Recent years have witnessed the hype of distant learning and replacement of dissection courses by plastic human models (ie, phantoms), particularly in the United States. However, for the aims outlined earlier (ie, "Introduction" section of this paper), training options for autonomous robots, such as simulations or phantoms, and even animal dissections, are still no match for training with human cadavers.

For gender, racial, or social status (eg, homeless patient) biases, we refer to two different aspects: (a) bias in the input data (eg, gender, race, or social status of the expert-in-the-loop[47]); and (b) biases in the training sets (eg, gender, race, or social status of the patient cadaver). One must ensure that there is no bias in both input data and training sets. For example, one must use cadavers that are fully representative of the full population. Machine learning is quite sensitive to the dataset bias. Therefore, we propose one can use digitalized anatomical atlases that are registered to the imaged cadaver using deformable registration. Hereby, contours, bones, and blood vessels could act as reference. Optimization of this process would again require AI. Moreover, one needs different types of reference models. If the training set is biased by gender or ethnicity in ways that correlate with anatomical differences, performance could be worse on those segments of the population that are not adequately represented in the training set.

For gender and "surgical ego" issues, it can be taken into consideration that AI robotic surgery can accomplish quite unwanted results. These unwanted results are (a) more gender inequality; and (b) more surgeons (male and female) self-focused on their own careers rather than focused on the well-being of the patients. Ego-driven attitudes, which may constitute a risk for patients, are enhanced by introducing sophisticated robotic technologies.[51] Their application may be supported by the disruptive behavior of the surgeons rather by therapeutic indications. Thus, along with the technological innovation, a cultural innovation is also needed in order to avoid such side effects.

For these specific problems of gender, racial, or social status biases, practical solutions need to be put in place. In order to guarantee the absence of bias, the learning data might be searchable. This implies that this data is kept in order to trace it. This retention of learning traces will allow auditing of data, especially in order to eliminate errors, inconsistencies, biases, and discrimination in the system. This data retention would play an active role in the event of a medical accident, in case the developer, the trainer, or the surgeon is blamed.

Attention should also be paid to the particular nature of personal data that may require compliance with ethical or legal rules. The European Union has, for example, put in place very strict regulations on the processing of personal data in 2016, especially with regard to health data.[53] Similar protection is found in some African countries, such as in Ivory Coast.[58] African Union members adopted in 2014 the African Union Convention on Cybersecurity and Personal Data Protection to develop guidelines on the protection of personal data.[59]

3.4 | Future of surgery

Objectively, one could aim to develop the surgical planning and skills for autonomous navigation through a human cadaver, and today, templates already exist for certain surgical and endovascular procedures. For this objective, the robotic system will learn how to (a) plan the steps to reach a desired location inside the human body; and (b) implement the skills required to execute the plan.

During the planning phase, the system will provide a sequence of dissection steps and decide which organs (tissue or group of tissues thereof), growth or foreign objects, etc, must be moved (or removed) to expose something specific, or alternatively, perform advanced needle biopsy with a simple robot so that whole organs/tissues do not need to be moved (or removed).

For autonomous surgery, tasks can include planning and executing procedures such as cutting, moving (or removing), sampling, and immunohistochemical tissue analysis. In the case of sampling and analysis without moving (or removing) organs/tissues, one may merely need to accurately place the needle in coordinates defined in a preinterventional image. The moving (or removing) of each organ/tissue will follow standard surgical plans to expose and access other organs/tissues. This is somewhat similar to classic AI planning (STRIPS and ADL programming), where a sequence of milestones must be achieved in a specific order to reach a goal state. The main difference here is that the application domain relates to moving (or removing) human organs/tissues instead of geometric figures. Although the best practices in work on cadavers will provide general guidelines, one could develop a system that will learn from surgeons on how to adjust the plan when facing unconventional cases by using more advanced learning algorithms.

Furthermore, the specific procedures for the removal of each organ will be implemented by studying and adapting surgeons' methods to the reality of the sensors and hardware. Much like the way STAR adapted anastomosis and electro-cautery techniques to its hardware and sensors, this proposed research will focus on implementing basic dissection skills that will be used in the context of moving (or removing) organs/tissues.

Like the STAR, this new research will rely on exogenous and endogenous fiducials, for example, external trackers, blood vessels, and bones. These fiducials in conjunction with 3D medical imaging data will facilitate the segmentation of relevant organs/tissues to be spared/removed or planning of the needle trajectory (one may use a flexible needle). Depending on the circumstances, one could simply use one arm or instead perform bimanual techniques by using two arms: one equipped with a forceps for manipulating tissues, and the other with an electro-cautery probe for cutting tissues.

The system can be evaluated on the successful harvesting of specific organs inside the body with little or no human assistance. For cadaver studies, we are not concerned with toxicity of agents/tracers (used with a perfusion device) or even radiation effects and dizziness from excessive medical imaging. However, hacking or the cybersecurity of these autonomous robots is still an important issue. We take into consideration that new developments for phantoms (such as synthetic human models, for example, SynDaver[60]) may also bring new opportunities for training autonomous surgical robots.

4 | CONCLUSION

The human anatomy is far more complex than textbook knowledge. Nevertheless, cadaver studies facilitate a "personalized" environment for the further development of the emerging field of "Explainable AI Robotics." We discussed specific examples of potential gender, racial, or social biases in AI, and in so doing, we articulated the benefits that transparency would provide for training autonomous surgical robots. However, robotics is a field that is only ideal for large study groups, and it is a field that may need some decades to be successful with machine learning approaches. One of the most commonly cited shortcomings of AI/machine learning in the medical robotics community is scarcity of data, or the cost and effort to generate it. Cadaver research addresses generating hundreds of datasets from surgical procedures with a surgeon-in-the-loop. These data can be made available to the community to develop novel methods and algorithms and benchmark results to existing ones, similar to the purpose of JIGSAWS for surgeon motion.[61]

We recommend "a procedure/skill template" for teaching AI that can be used by a surgeon. Similar existing methodologies show that when this metric-based approach is used for training surgeons, cardiologists, and anesthetists, it results in a >40% error reduction in objectively assessed intraoperative procedures.[28-30,42-45,62-64] Nowadays, authors reviewing surgical robotics are putting much emphasis on comparing surgical skills of "humans" vs "robots." We submit that cadaver studies that involve morphological and molecular imaging can be used to advance the ability to objectively evaluate the results of autonomous robotic performance. Students learn their skills by dissecting cadavers. The state of this art can be greatly enhanced through the use of today's ground-breaking technologies by “training” autonomous robots to undertake such dissections with superior skill. It can be extrapolated that, in time, cadavers can be substituted by living patients.

While we envision autonomous robotic “surgery” as a long-term goal. It is not tabled as an immediate solution with practical implications. The expert-in-the-loop[65] emphasizes that computers should do what they can do good, and that humans should do what they can do good, so augmenting human intelligence with artificial intelligence (AI)—but not replacing humans by AI—is important because, particularly in medicine (and in many other domains), the ultimate responsibility lies with the human (eg, surgeon). This requires that the human-in-the-loop is able—ondemand—tore-trace to re-enact and to understand the machine decisions, having a chance to interact with the AI, which needs effective human-AI interactions and measurements for causal reasoning.[66]

We do acknowledge that there are some limits of surgical training on cadavers including, for example, that the skills required to work with a living body, such as managing bleeding, working around pulsing arteries, etc, are not present in the cadaver. In itself, the removal of organs/tissues will require a systematic plan for dissecting arteries, veins, and bronchus. Planning algorithms will combine surgeon expertise with the constraints imposed by mechanical and sensing limitations. For the execution phase, the robot may implement skills for incisions, exposing, and moving (or removing) organs/tissues, growth, or foreign objects, etc. Successful execution may rely on using tracing agents to highlight the areas of interests and guide the motion of the robot. Methods, devices, or systems, tested with cadavers, will also be easier to translate to in vivo surgeries and medical robotics in general—the second fastest growing market in robotics and is predicted to increase 5-fold over the next 3 years.[67]

REFERENCES

1. Jeelani S, Dany A, Anand B, Vandana S, Maheswaran T, Rajkumar E. Robotics and medicine: a scientific rainbow in hospital. *J Pharm Bioallied Sci*. 2015;7(suppl S2):381-383.
2. Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int J Med Robot*. 2011;7(4):375-392. https://doi.org/10.1002/rcs.408.
3. Holzinger A, Kickmeier-Rust MD, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. *Comput Educ*. 2009;52(2):292-301. https://doi.org/10.1016/j.compedu.2008.08.008.
4. Pabst R. Gross anatomy: an outdated subject or an essential part of a modern medical curriculum? Results of a questionnaire circulated to final-year medical students. *Anat Rec*. 1993;237:431-433.
5. Cardell RR. The education and training of anatomy graduate students in the biomedical sciences in the twenty-first century. *Anat Rec*. 1998;253:41-41.
6. Eizenberg N. Anatomy and its impact on medicine: will it continue? *Australas Med J*. 2017;8:373-377. https://doi.org/10.4066/AMJ.2015.2550.
7. McCall M. The secret lives of cadavers. National Geographic. https://news.nationalgeographic.com/2016/07/body-donation-cadavers-anatomy-medical-education/. Published July 29, 2016. Accessed 4 November 2019.
8. United States Department of Defense. (2017). DoDFY17 Medical Simulation and Information Sciences FORwARD Award cdmrp.army.mil/funding/pa/FY17-DMRDP-MSISRP-FOR.pdf. Accessed 4 November 2019.
9. United States FDA. (2018). Software Precertification Program: Working Model—Version 0.2. 2018 Available from: http://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM611103.pdf. Accessed 10 October 2018.
10. 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*. 2018;2018(1):e1968. https://doi.org/10.1002/rcs.1968.
11. Ficuciello F, Tamburrini G, Arezzo A, et al. Autonomy in surgical robots and its meaningful human control. *Paladyn, J Behav Robot*.2019;10(1):30-43. https://doi.org/10.1515/pjbr-2019-0002.
12. Fichera L, Dillon NP, Zhang D, et al. Through the eustachian tube and beyond: a new miniature robotic endoscope to see into the middle ear. *IEEE Robot Autom Lett*. 2017;2(3):1488-1494. https://doi.org/10.1109/LRA.2017.2668468.
13. Yang S, MacLachlan RA, Martel JN, Lobes LAJr, Riviere CN. Comparative evaluation of handheld robot-aided intraocular laser surgery. *IEEE Trans Robot*. 2016;32(1):246-251.
14. Fornalik H, Fornalik N, Kincy T. Advanced robotics: removal of a 25 cm pelvic mass. *J Minim Invasive Gynecol*. 2015;22(6S):S154. https://doi.org/10.1016/j.jmig.2015.08.575.
15. Tsai TY, Dimitriou D, Li JS, Kwon YM. Does haptic robot-assisted total hip arthroplasty better restore native acetabular and femoral anatomy?*Int J Med Robot*. 2016;12(2):288-295. https://doi.org/10.1002/rcs.1663.
16. Feng AL, Razavi CR, Lakshminarayanan P, et al. The robotic ENT microsurgery system: a novel robotic platform for microvascular surgery. *Laryngoscope*. 2017;127(11):2495-2500. https://doi.org/10.1002/lary.26667.
17. Ashrafian H, Clancy O, Grover V, Darzi A. The evolution of robotic surgery: surgical and anaesthetic aspects. *Br J Anaesth*. 2017;119(1):i72-i84.
18. Shademan A, Decker R, Opfermann JD, Leonard S, Krieger A, Kim PC. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med*. 2016;8(337):337ra64. https://doi.org/10.1126/scitranslmed.aad9398.
19. Leonard S, Wu KL, Kim Y, Krieger A, Kim PCW. Smart tissue anastomosis robot (STAR): a vision-guided robotics system for laparoscopic suturing. *IEEE Trans Biomed Eng*. 2014;61(4):1305-1317.
20. Opfermann JD, Leonard S, Decker RS, et al. Semi-autonomous electrosurgery for tumor resection using a multi-degree of freedom electrosurgical tool and visual servoing. Paper presented at: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); September24-28, 2017; Vancouver, Canada: 3653-3659. https://doi.org/10.1109/IROS.2017.8206210.
21. Helen L, et al. “Investigation of tissue bioimpedance using a macro-needle with a potential application in determination of needle-to-nerve proximity.” International Journal on Smart Sensing and Intelligent Systems, 2019; 12(2).
22. Illanes A, Esmaeili N, Poudel P, Balakrishnan S, Friebe M. Parametrical modelling for texture characterization—A novel approach applied to ultrasound thyroid segmentation. *PLoS ONE*. 2019;14(1):e0211215. https://doi.org/10.1371/journal.pone.0211215.
23. Illanes A, Boese A, Maldonado I, et al. Novel clinical device tracking and tissue characterization using proximally placed audio signal acquisition and processing. *Sci Rep*. 2018;8(1):12070. https://doi.org/10.1038/s41598-018-30641-0.
24. Chen CH, Suehn T, Illanes A, et al. Proximally placed signal acquisition sensoric for robotic tissue tool interactions. *Curr Dir Biomed Eng*. 2018;4(1):67-70.

25. van Oosterom MN, van der Poel HG, Navab N. Computer-assisted surgery: virtual- and augmented-reality displays for navigation during urological interventions. *Curr Opin Urol*. 2018;28(2):205-213.
26. van den Berg NS, Buckle T, KleinJan GH, van der Poel HG, van Leeuwen FW. Multispectral fluorescence imaging during robot-assisted laparoscopic sentinel node biopsy: a first step towards a fluorescence-based anatomic roadmap. *Eur Urol*. 2017;72(1):110-117.
27. Decker RS, Shademan A, Opfermann JD, Leonard S, Kim PC, Krieger A. Biocompatible nearinfrared three-dimensional tracking system. *IEEE Trans Biomed Eng*. 2017;64(3):549-556.
28. Van Sickle KR, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg*. 2008;207(4):560-568. https://doi.org/10.1016/j.jamcollsurg.2008.05.007.
29. Pedowitz RA, Nicandri GT, Angelo RL, Ryu RKN, Gallagher AG. Objective assessment of knot-tying proficiency with the fundamentals of arthroscopic surgery training program workstation and knot tester. *Arthroscopy*. 2015;31(10):1872-1879.2.
30. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann surg*. 2005;241(2):364-372.2.
31. Strickland S. Fukushima's next 40 years. *IEEE Spectrum*. 2014;51(3): 46-53.
32. Thrun S, Beetz M, Bennewitz M, et al. Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *Int J Rob Res*. 2000;19(11):972-999
33. Vallor S, Bekey GA. In: Lin P, Abney K, Jenkins R, eds. *Robot Ethics 2.0 Artificial Intelligence and the Ethics of Self-learning Robots*. New York: Oxford University Press; 2017:338-353.
34. Bolukbasi T, Chang KW, Zou JY, Saligrama V, Kalai AT. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Paper presented at: 30th Conference on Neural Information Processing Systems (NIPS 2016): Barcelona, Spain. 4349-4357.
35. McDuff D, Cheng R, Kapoor A. *Identifying bias in AI using simulation*. arXiv:1810.00471v1 [cs. LG]. September 30, 2018.
36. Fingerhut A. Surgical ego, the good, the bad, and the ugly. *Surg Innov*.2011;18(2):97-98.
37. Myers CG, Lu-Myers Y, Ghaferi AA. Excising the “surgeon ego” to accelerate progress in the culture of surgery. *BMJ*. 2018;363:k4537. https://doi.org/10.1136/bmj.k4537.
38. Kirkpatrick T, LaGrange C. Robotic surgery: risks vs rewards. *AORN J*. 2017;106(2):186.
39. Cooper MA, Ibrahim A, Lyu H, Makary MA. Underreporting of robotic surgery complications. *J Healthc Qual*. 2015;37(2):133-138. https://doi.org/10.1111/jhq.12036, https://doi.org/10.1111/jhq.12036.
40. Ahmed F, Rhee J, Sutherland D, Benjamin C, Engel J, Frazier HII. Surgical complications after robot-assisted laparoscopic radical prostatectomy: the initial 1000 cases stratified by the clavien classification system. *J Endourol*. 2012;26(2):135-139.
41. Gallagher A. Metric-based simulation training to proficiency in medical education:-What it is and how to do it. *Ulster Med J*. 2012;81(3): 107-113.2.
42. Srinivasan KK, Gallagher A, O'Brien N, et al. Proficiency-based progression training: an ‘end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomised control study. *BMJ Open*. 2018;8(10):e020099.3.
43. Cates CU, Lönn L, Gallagher AG. Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators. *BMJ Simul Technol Enhanced Learn*. 2016;2(1):1-5.
44. Angelo RL, Ryu RK, Pedowitz RA, et al. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy*. 2015;31(10):1854-1871.5.
45. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg*. 2007;193(6): 797-804.
46. RoboLaw. Guidelines on regulating robotics. http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf. 2014. Accessed 10 November 2019.
47. Holzinger A, Plass M, Holzinger K, Crisan GC, Pintea CM, Palade V. *A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop*. arXiv:1708.01104 [cs.AI] (or arXiv:1708.01104v1 [cs.AI]). August 3, 2017.
48. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639): 115-118. https://doi.org/10.1038/nature21056.
49. Humphreys P. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press; 2004.
50. Nevejans N. *The Influence of Decision Support Software on Medical Decision Making in the Light of Law and Ethics*. *Innovations en santé publique, des données personnelles aux données massives* (in Paris: French). Dalloz; 2018:113-127.
51. Boscarato C. Robotics, innovation and the law. *The Challenge of Innovation in Law*. Vol 221. Pavia, Italy: Pavia University Press; 2015.
52. Nevejans N. *Traité de droit et d éthique de la robotique civile*. Treatise of Law and Ethics of Civil Robotics (in English). LEH ed. 2016;646-651:52.
53. Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. L 119/1. 4.5.2016.
54. Schölkopf B, Platt J, Hofmann T. *Robotic Grasping of Novel Objects. Advances in Neural Information Processing Systems*. Boston, MA: MIT; 2007:1209-1216.
55. Siegwart R, Nourbakhsh IR, Scaramuzza D. *Introduction to Autonomous Mobile Robots*. Boston, MA: MIT Press; 2011.
56. Knox WB, Stone P, Breazeal C. Training a robot via human feedback: a case study. In: Herrmann G, Pearson M, Lenz A, Bremner P, Spiers A, Leonards U, eds. *Social Robotics*. Lecture Notes in Artificial Intelligence (LNAI). Vol 8239. Berlin, Germany: Springer; 2013;2013: 460-2470.
57. Franz MO, Schölkopf B, Mallot HA, Buelthoff HH. Learning view graphs for robot navigation. *Auton Rob*. 1998;5(1):111-125.
58. Law No. 2013-450 dated June 19, 2013 on the protection of personal data. http://www.ictpolicyafrica.org/fr/document/4wo0y6uby6j?page=1.
59. African Union convention on cyber security and personal data protection, June 27, 2014. https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection.
60. Weisberger M. (2017). *It Bleeds. It Breathes. It's a Lifelike Artificial Human Corpse!* New York: Live Science publisher. https://www.livescience.com/61089-syndaver-synthetic-humans.html.
61. Ahmidi N, Tao L, Sefati S, et al. A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. *IEEE Trans Biomed Eng*. 2017;64(9):2025-2041. https://doi.org/10.1109/TBME.2016.2647680.
62. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):458-463; discussion 463-454.
63. Breen D, O'Brien S, McCarthy N, Gallagher AG, Walshe N. Effect of proficiency based progression simulation training and standard simulation training on ISBAR performance. a randomized controlled trial. *BMJ Open*. 2019;9e025992. https://doi.org/10.1136/bmjopen-2018-025992.

64. Gallagher AG, O'Sullivan GC. *Fundamentals of Surgical Simulation: Principles and Practices*. London, UK: Springer Verlag; 2011.
65. Holzinger A, Plass M, Kickmeier-Rust M, et al. Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Appl Intell*. 2018;49:2401-2414. https://doi.org/10.1007/s10489-018-1361-5.
66. Holzinger A. From machine learning to explainable AI. Paper presented at: 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA); August23-25, 2018; Kosice, Slovakia: 55-66. https://doi.org/10.1109/DISA.2018.8490530.
67. Litzenberger G, et al. Welcome to the IFR Press Conference. IFR. Tokyo. 2018. https://ifr.org/downloads/press2018/WR_Presentation_Industry_and_Service_Robots_rev_5_12_18.pdf Last accessed 15 October 2019.

How to cite this article: O'Sullivan S, Leonard S, Holzinger A, et al. Operational framework and training standard requirements for AI-empowered robotic surgery. *Int J Med Robotics Comput Assist Surg*. 2020;16:e2020. https://doi.org/10.1002/rcs.2020