

Ethical implications of AI in robotic surgical training: A Delphi consensus statement

Authors: Justin W. Collins[1,2,3*], Hani J Marcus[2], Ahmed Ghazi[4], Ashwin Sridhar[1,3], Daniel Hashimoto[5], Gregory Hager[6], Alberto Arezzo[7], Pierre Jannin[8], Lena Maier-Hein[9], Keno Marz[9], Pietro Valdastri[10], Kensaku Mori[11], Daniel Elson[12], Stamatia Giannarou[12], Mark Slack[13,14], Luke Hares[15], Yanick Beaulieu[16], Jeff Levy[17], Guy Laplante[18], Arvind Ramadorai[19], Anthony Jarc[20], Ben Andrews[21], Pablo Garcia[22], Huzefa Neemuchwala[23], Alina Andrusaite[22], Tom Kimpe[24], David Hawkes[2], John D. Kelly[1,2,3], Danail Stoyanov[2]

Affiliations: 1. University College London, Division of Surgery and Interventional Science, Research Department of Targeted Intervention; 2. Wellcome/ESPRC Centre for Interventional and Surgical Sciences (WEISS), University College London; 3. University College London Hospital, Division of Uro-oncology; 4. Simulation Innovation Laboratory, University of Rochester, USA; 5. Surgical Artificial Intelligence and Innovation Laboratory, Massachusetts General Hospital, USA; 6. Malone Center for engineering in healthcare, Department of Computer Science, John Hopkins University, Baltimore, USA; 7. Department of Surgical Sciences, University of Torino, Italy; 8. University Rennes, Inserm, Rennes, France; 9. Deutsches Krebsforschungszentrum, Division of Computer Assisted Medical Interventions, Heidelberg Germany; 10. STORM Lab, School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK; 11. Director of Information Technology Center, Nagoya University, Japan; 12. Hamlyn Centre for robotic surgery, Department of Surgery and cancer, Imperial College London, UK; 13. Honorary Senior Lecturer, University of Cambridge, Cambridge UK; 14. CMO CMR Surgical, Cambridge, UK; 15. Chief technology director, CMR Surgical, Cambridge, UK; 16. Division of Cardiology and Critical Care, Sacré-Coeur Hospital, University of Montreal, Montreal, Canada; 17. Institute for Surgical Excellence, Philadelphia, USA; 18. Director, Global Medical Affairs at *Medtronic* Minimally Invasive Therapies, Brampton, Canada; 19. Director, Digital-Assisted Surgery (DAS), Medtronic Surgical Robotics, North Haven, CT, USA; 20. Applied Research, Intuitive Surgical, Inc., Sunnyvale, CA, USA; 21. Strategy, Intuitive Surgical, Inc., Sunnyvale, CA, USA; 22. Johnson and Johnson Medical Devices, CA, USA; 23. C-SATS Inc., a Johnson & Johnson Medical Devices Company, CA, USA; 24. BARCO NV – Healthcare division, Kortrijk, Belgium

Introduction:

Healthcare is perceived as a biomedical industry, but it has also always been an information industry and with an exponentially growing number of data collection devices there are increasing opportunities for machine learning (ML) and artificial intelligence (AI) applications. Healthcare is one of the most promising areas for the integration of AI systems, but it is also one of the most complex. The healthcare sector incorporates many regulatory, privacy and ethical laws that aim to ensure innovation will 'first cause no harm'. As the role of AI in healthcare continues to expand there is increasing awareness of the potential pitfalls of AI and the need for guidance to avoid them.

Objectives:

To provide ethical guidance on developing narrow AI applications for surgical training curricula. We define standardised approaches to developing AI driven applications in surgical training that address current recognised ethical implications of utilising AI on surgical data. We aim to describe an ethical approach based on the current evidence, understanding of AI and available technologies, by seeking consensus from an expert committee.

Materials and Methods:

The project was carried out in 3 phases: (1) A steering group was formed to review the literature and summarize current evidence.

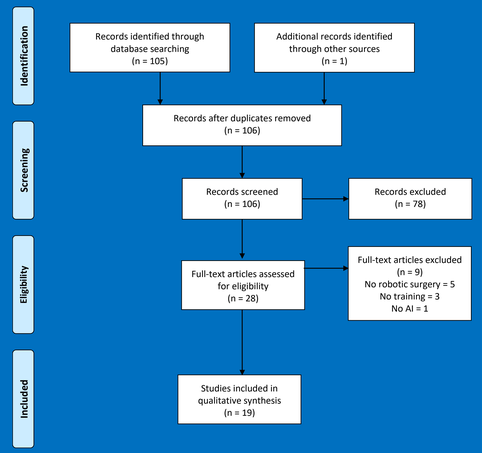


Figure 1. PRISMA flow diagram summarising the study selection process.

Materials and Methods (continued):

(2) A larger expert panel convened and discussed the ethical implications of AI application based on the current evidence. A survey was created, with input from panel members. (3) Thirdly, panel-based consensus findings were determined using an online Delphi process to formulate guidance. 30 experts in AI implementation and/or training including clinicians, academics and industry contributed. The Delphi process underwent 3 rounds. Additions to the second and third-round surveys were formulated based on the answers and comments from previous rounds. Consensus opinion was defined as ≥ 80% agreement.

Results:

There was 100% response from all 3 rounds. The resulting formulated guidance showed good internal consistency, with a Cronbach alpha of >0.8. There was 100% consensus that there is currently a lack of guidance on the utilisation of AI in the setting of robotic surgical training. Consensus was reached in multiple areas, including: 1. Data protection and privacy; 2. Reproducibility and transparency; 3. Predictive analytics; 4. Inherent biases; 5. Areas of training most likely to benefit from AI.

Level of agreement	General statements about the potential for telepresence in robotic surgery training
100%	• There are potential benefits to the utilisation of AI in the setting of surgical training curricula • There is currently a lack of guidelines (or guidance) on the utilisation of AI in the setting of surgical training • The future success of AI in surgical training will require its ethical deployment within healthcare organisations
≥95%	• There are potential risks to the utilisation of AI in the setting of surgical training curricula • An aim of this group is to identify the ethical implications and to formulate guidance on AI in surgical training • A lack of education around the use of AI can impact misuse, misinterpretation, misapplication of AI techniques and results

Table 1: General questions on the ethical implications of AI in surgical training .

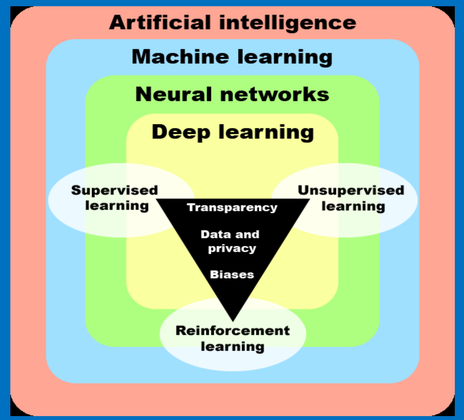


Figure 2: Relationship of terminology and link to ethical implications.

Term	Definition
General artificial intelligence (GAI)	Artificial general intelligence (AGI) is the hypothetical intelligence of a machine that has the capacity to understand or learn any intellectual task that a human being can.
Narrow AI	A term used to describe artificial intelligence systems that are specified to handle a singular or limited task.
Natural language processing (NLP)	The relationship between computers and human language. Natural language refers to speech analysis in both audible speech and text. NLP systems capture meaning from an input of words (sentences, paragraphs, pages, etc.)
Algorithm	An algorithm is a formula representing the relationship between certain variables. An algorithms is a set of instructions with a finite end designed to produce an output.
Learning algorithms	A method used to process data to extract patterns appropriate for application in a new situation. In particular, the goal is to adapt a system to a specific input-output transformation task.
Cluster	Any group of data, people or things that share a common characteristic. Data is combed through by AI to uncover patterns and connections that humans might not readily see.
Machine learning	With limited need for programming, machines can teach themselves. Combing through massive amounts of data allows AI to identify patterns and groupings.
Model	A trained convolutional neural network (CNN) or other neural network, where the output of a training phase and the coefficients of an AI (or ML) model are used to perform inference – run on new data and make predictions
Data sets	Data sets are divided into a training set (with labels provided) for learning and a test set (no labels provided) that allows for the assessment of the performance of the algorithm on new data.
Training datasets	The training dataset is an initial set of data used to help a program understand how to apply ML technologies like neural networks to learn and produce sophisticated results. Training dataset is also known as a training set, training data or learning set.
Robotic surgery training data (RSTD)	Data collected from robotic surgery devices that relates to surgeon's performance. Data may be collected and recorded manually or with automated mechanisms
Deep learning	Deep learning is a subset of ML where artificial neural networks, algorithms inspired by the basic structure of the human brain, learn from large amounts of data. Deep learning enables machines to solve complex problems relying on the underlying data used to train a machine, rather than on human problem modelling..
Image understating / Computer vision	Sub-branch of AI focused on the development of algorithms to infer information from images or video, mimicking human cognitive sense based on the eyes.
Supervised learning	ML that requires human input. Supervised learning is a task-driven process wherein an algorithm is trained to predict a prespecified output, such as identifying a stop sign in a photograph. The "supervised" moniker comes from the need to provide annotated (ie, labeled) data so that it can learn the associations between inputs and the desired output.
Unsupervised learning	Requiring no human labels in the training data or limited (semi-supervised), this type of learning allows Alto draw conclusions on its own from patterns that it finds.
Reinforcement learning	ML where learning occurs through successive attempts via trial and error and rewards/punishments guide the behavior of the model to optimize rewards. Fitting to problems where complex games/conditions can be exectured many times in scenarios that explore the full problem domain. A famous example is the Google Deepmind AlphaGoZero algorithm.
Outcome measures	Quantifiable consequences of an action, set of actions, or procedure; also the final result that is to be measured (time, speed, accuracy, performance errors, leadership, communication skill)
Metric	Operational definition of an entity, object, location, task, error etc.; must support an outcome measure ("no measurement, no metric")
Score	Number of points, goals, etc. achieved during performance
Benchmark	A defined point of reference against which things may be compared; it is to be noted that this is not the score
Standard	The authorized exemplar of a unit of measure (performance metric: an agreed "external" metric, usually defined by an authorised body)
Hazard step	A step in a procedure that is associated with commonly occurring or recognised surgical errors
Risk prediction	A risk prediction model is an algorithm that uses patient risk factor data to estimate the probability of a patient experiencing a healthcare outcome. In surgery, they are commonly used to predict the risk of adverse outcomes after intervention.
Early warning system	A technology, ML algorithm or associated policies and procedures designed to predict and mitigate patient harm and other undesirable events
Near miss	An unplanned event that had the potential to cause harm, but did not actually result in human injury, or equipment damage or an interruption to normal operation
Data minimisation	Collecting the minimum amount of personal data that you need to deliver an individual element of your service or to enable machine learning algorithms
Data normalisation	Attempting to normalise data to some commonly agreed criteria
Telemetry	Automated performance metrics (APMs) generated by the robotic device related to kinematic data
Transparency	The ability of the person handling AI output data, to determine how an algorithm reaches its conclusion
Confirmation bias	Occurs when data/answers are used to confirm hypothesis or beliefs
Interpretation bias	Occurs when researchers' conclusions or assessments are affected by descriptions e.g. what speed was the car going when it hit the fence, compared with what speed was the car doing when it smashed into the fence
Prediction bias	Occurs when data points are too focused so that association is concluded. Can result from poor programing in AI
Information bias	Occurs when researchers or programmers use information or data, they believe is already linked to their outcome

Table 2 – Clarifications on terminology.

General guidance for anonymising data to protect privacy
1. Everyone handling the data has shared responsibility to deidentify data
2. The data protection officer (DPO) has overall responsibility for data protection compliance matters
3. All data that is planned to be collected, to be proactively approved, and stored according to guidelines from the organisation's data protection office
4. The responsibilities of a DPO include: • **Informing and advising organisations of their obligations under data protection law** • **Monitoring compliance with the regulations and related policies, including raising of awareness and training of staff** • **Providing procedures, guidance and advice in support of this policy e.g. for Data Protection Impact Assessment (DPIAs)** • **Acting as the organisations first point of contact with the Information Commissioner's Office (ICO)** • **Handling subject access requests and official requests for personal data from third parties** • **Investigating losses and unauthorised disclosures of personal data**
5. Organisational accountability under GDPR, is legally obliged to put into place: • **Comprehensive governance issues** • **Privacy impact assessments** • **Privacy measures by design**
6. Lawfulness: processing data has to be done for a specific purpose to which the user has agreed and has to match up with how it is described
7. Purpose limitations: data is to be used for a specific purpose of which the user has been made aware through explicit consent
8. Data minimisation: review what data you have and why. Only capture the minimum amount of data you need
9. Data accuracy: make sure that the data is accurate and ideally stored in a way that allows the user to update or delete the data themselves (securely)
10. Storage limitations: data that is no longer required should be removed. If kept for longer than needed data should be pseudonymised to protect users identity
11. Integrity: processors should protect user data against unlawful processing or loss. Ideally having encryption of user data and privacy by design processes

Table 3: General guidance for anonymising data to protect privacy.

Element	Description
Phase	A section of the procedure, with a clearly defined start and end point. Phases of a procedure enable modular training, with training commenced in the easier phases with a lower frequency of hazard steps
Visual cue	A visual cue is an anatomical landmark or areas of interest that is defined as important and therefore needs to be identified within a given phase of the procedure. Image recognition ML that identify visual cues, need to be able to label this data consistently and accurately
Task	Task Defined steps to be completed within a phase of the procedure. Tasks can be further deconstructed into manoeuvres and surgical gestures [20].
Technique error	An error of the technique that may or may not be associated with an event (eg, using a wrong instrument to grasp a bowel). Technique errors may be associated with near misses and have the potential to enable early warning systems
Event error	An error of the technique or a device error that results in a harm to tissue, patient, staff or equipment etc
Device error	Device malfunction or failure
Telemetry	Automated performance metrics generated by the robotic device related to kinematic data
Tissue characteristics	Agreed terms for tissue related to ease of operation. For example inflamed, fibrotic etc
Anatomical variations	Identified common anatomical variations that may impact ease of the operation

Table 4: Elements of surgical procedure to be identified for data normalisation .

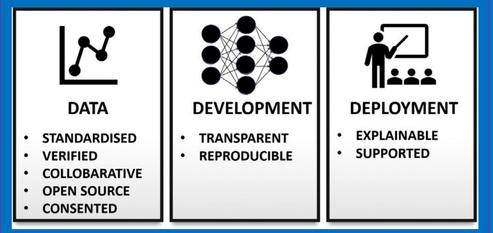


Fig. 3 – Summary of the guidance for optimising the application of AI in robotic surgery training..

Conclusions:

Using the Delphi methodology, we achieved international consensus among experts to develop and reach content validation for guidance on ethical implications of AI in surgical training. Providing an ethical foundation for launching narrow AI applications in surgical training. This guidance will require further validation.

Funding The work was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) UK [203145Z/ 16/Z]; Engineering and Physical Sciences Research Council (EPSRC) UK [EP/P027938/1, EP/R004080/1, EP/P012841/1]; The Royal Academy of Engineering Chair in Emerging Technologies scheme, UK.

Acknowledgements: We would like to thank David Hayward, a BBC journalist, lecturer on ethics and patient advocate who contributed to the meetings and Delphi process.

SAGES 2021 Aug 31 - Sept 3, 2021 MGM Grand, Las Vegas, Nevada

@4urology