

When and what patients need to know about AI in clinical care

David Shaw^{ab}, Giorgia Lorenzini^a, Laura Arbelaez Ossa^a, Jens Eckstein^c, Luzius Steiner^c, Bernice S. Elger^{ad}

^a Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

^b Care and Public Health Research Institute, Maastricht University, Maastricht, the Netherlands

^c Basel University Hospital, Basel, Switzerland

^d Center for Legal Medicine, University of Geneva, Geneva, Switzerland

Summary

In this article, we describe and analyse when and what patients need to be told about the use of artificial intelligence in clinical care. In many circumstances, patients do not need to be told about the use of AI, but in cases in which AI is making decisions rather than supporting them or patient data are being transferred to a third party, specific consent must be sought from patients. Ultimately, AI will become a routine part of clinical care, but until then, clinicians must take great care to seek consent where necessary.

All doctors know that patients must give consent to clinical care because they have the right to make decisions about their bodies and data. However, the introduction of artificial intelligence (AI) to diagnostic and treatment pathways complicates matters, raising important questions. Should patients consent to the use of AI? Do they just need to be informed about its use? Or is that going too far, and should AI simply be considered another healthcare tool? In this article, we aim to answer these questions from both an ethical and a legal perspective to enable doctors to use AI safely and confidently with patients. Our discussion is based on empirical research (qualitative interviews) involving doctors and experts working with AI [1, 2], as well as legal analysis conducted with lawyers [3], as part of the Swiss National Science Foundation NRP77 project EXPLAIN. Although our legal analysis is specific to Switzerland, our conclusions are broadly transferable to other jurisdictions in Europe, given the similar ethico-legal norms in clinical practice and data protection.

Whether patients know it or not, AI is already used to advance clinical care, and there have been recent calls for improving regulation in the medical field [4]. AI has been shown to increase the accuracy of breast cancer screening [5], and it assists surgeons during operations [6]; it also reduces workload across a wide range of specialities and administrative tasks [7]. In addition, it can assist doctors by performing depth of anaesthesia monitoring [8], and it is even used in automated defibrillators [9]. AI has also shown potential across other areas of medicine, including global health [10], genomics [11], and bioethics [12], but our focus in this paper is on current and upcoming uses in

clinical care, in which the potential advantages for patients are already clear.

Extensive literature already addresses the ethical issues raised by AI, including the need for explainability, risks of discrimination (for example, AI is more likely to miss skin cancer in people with darker skin [13]), threats to confidentiality, and even unreliable chatbots [14, 15]. However, one issue has remained relatively neglected: although the use of AI in healthcare is increasing rapidly, many patients remain unaware of this. Given the importance of informed consent to both patient autonomy and the relationship between doctors and patients, should doctors seek explicit consent to the use of AI, or at least inform patients about its use?

In answering this question, it is important to remember the purpose of seeking consent: to respect the patient's autonomy by allowing them to make informed decisions about their care [16]. This cannot be accomplished through their agreement to the use of something that they do not fully understand. If a doctor simply says, "We're using AI as part of your care, is that ok?" without explaining what this means, patients might agree, but it would not be an informed choice. Therefore, if doctors are to seek consent, they must be clear about exactly what they are seeking consent for to facilitate shared decision-making [17].

Consent in medicine also plays a role in enabling respect for patients' autonomy in balancing risks and benefits. In several situations (e.g. when it comes to balancing length and quality of life), rational persons might disagree in their evaluation of which alternative is better. The same applies to the use of AI. Table 1 provides current clinical care examples in which consent to use AI is necessary and examples of when it is not.

In many cases, AI is used as a clinical decision support system, such as in depth of anaesthesia monitoring. Doctors can use the outputs of such systems to inform their own evaluation of a patient's care. In such cases, AI helps doctors by providing extra information that can benefit patients. As such, decision support systems are quite similar to other software and devices used regularly by doctors to inform their decisions and those of their patients. While AI might be a different type of technology, consent is not always sought to use specific software or devices, although patients might be informed of their use. In most cases, the

Dr David Shaw
Institute for Biomedical
Ethics
University of Basel
Bernoullistrasse 28
CH-4056 Basel
david.shaw[at]unibas.ch

same applies to the use of AI decision support systems: there is no ethical or legal obligation to mention their use, but doctors might want to do so in the spirit of full disclosure. If they do discuss AI with patients, they should be prepared to explain in more detail and answer questions so that patients are sufficiently informed.

However, there are two important exceptions to the general rule that patients do not need to be informed about the use of AI. First, although most AI systems are mainly advisory, if an AI system is being used not as a decision support system but as a decision-making system (in which decisions or recommendations are fully automated), then patients must be informed about this and should be able to consent to or refuse the use of such systems. Here, we do not refer to a simple “if A, then do B” logic but a process in which multiple parameters are interpreted and integrated to draw a conclusion in a manner comparable to human clinical reasoning. One example is in imaging, for which AI systems can be more accurate and reliable than doctors in classifying tumours and recommending treatment. Again, if doctors can interpret the results of the AI system and make their own recommendation, consent may not be necessary, but if the AI is independently recommending treatment, specific consent must be sought (even if the AI algorithm is fully “explainable”, meaning that the doctor understands why it is making the recommendation – see the next paragraph). Traditionally, doctors have made decisions in partnership with patients, with patient consent required to legally permit actions that would otherwise be considered bodily injury. If AI replaces the traditional decision-making process, this is not what a patient would implicitly expect. Therefore, if AI largely replaces the physician’s decision-making role, informing the patient will be necessary to ensure a true AI–patient partnership in decision-making.

Seeking consent from patients is particularly important in any situation in which doctors do not fully understand why AI systems are making decisions or recommendations (the so-called “black box” problem) (although the use of “unexplainable” AI of this type is currently unlikely in the healthcare context) [21]. Depending on how closely integrated AI is in care systems, this could pose practical challenges if patients do not want to use it, but their wishes should be respected when possible. If informed consent to the use of AI is sought formally, patients must be informed about the potential effects of its use on their care compared

to the consequences of not using AI, including the prospective risks and benefits of both alternatives, without overburdening them with complicated technical information.

While doctors also make decisions that are not fully or transparently explainable, introducing new technology to the clinical setting can necessitate holding it to a higher standard in order to build and maintain trust [22].

Second, while it might not be necessary to seek consent to the use of AI in terms of medical ethics, it might be necessary because of data protection laws. Patients must be informed about the purposes for which their data are used, but if AI is used like other hospital software, this will be covered by the general consent to treatment and data use within the hospital. However, if a hospital is sending patient data out of the hospital to a provider of AI services, then patients may have a right to be informed about this and opt out of such data transfer, depending on the local legislation (the same may apply if AI is used for any “high-risk” profiling of patients by algorithms designed for this purpose). However, if data are processed within the hospital, then the fact that the software involves AI does not in itself necessitate that patients be told of that particular processing of their data. Note that this requirement to disclose or seek consent to data transfer arises from data protection requirements, not from anything specific to the use of AI. (Another related potential need to seek consent stems from any potential commercial (re-)use of patients’ data stemming from sharing with a local or off-site third-party provider.)

Patient consent may also be needed when evidence shows that AI is generally better at determining the best treatment for a patient, but the chosen treatment poses significantly higher risks and the doctors disagree (for example, a chemotherapy that has particularly high risks). In such scenarios, any shared decision-making regarding treatment would clearly need to include a discussion of AI’s role.

A final important point to consider is that patients should perhaps be informed when AI could provide a higher standard of medical care but is not being used. If, as the evidence suggests, AI improves efficiency and outcomes, there may be a stronger case for informing patients when hospitals are *not* using AI systems, as their care is more likely to be compromised [23]. Generally, patients should be warned about relevant risks, and in future, not using AI may significantly increase risk; this in turn may impose an

Table 1:
Current uses of AI in medicine and where consent is necessary.

Examples of AI use	Explicit consent is necessary	Implicit consent is sufficient
Anaesthesiology	Closed loop systems dosing anaesthesia based on depth of anaesthesia monitoring [18] (as of today, not commercially available, experimental only) Consent is required because AI determines the dose	Depth of anaesthesia monitoring e.g. BIS™, Narcotrend™.
Anaesthesiology	Sedasys™ (Johnson & Johnson Robot for Sedation, e.g. for gastroenterological procedures) [19] Consent is required because AI determines the dose	Intraoperative pain monitoring (e.g. NOL™ pain response monitor); Acumen HPI software (Edwards) for the prediction of intraoperative hypotension up to 10 minutes prior to a hypotensive event
Internal medicine	Sinus rhythm ECG based prediction of atrial fibrillation in the future [20] Consent is required because the physician has never seen an ECG of the patient showing atrial fibrillation and must rely on the prediction of the AI with no possibility to verify it (In case of predicted atrial fibrillation, lifelong anticoagulation therapy should be given, which is associated with an increased risk of bleeding)	Drug interaction check (e.g. MEONA®)
Internal medicine	Conversation with a chatbot instead of a real person to take the history Consent is required to ensure that the patient knows that it is an AI, not a doctor.	Decision support tools to challenge the suspected diagnosis (Ada Health®)

ethical obligation to use AI whenever evidence indicates that it improves the quality of care. (In research settings, in which AI is not yet validated, consent should of course be sought from patients.)

Therefore, considering the analysis above, the answer to our original question is as follows: generally, formal patient consent is not required for the use of AI in clinical decision support systems, but doctors may choose to mention it informally, as they would with other tools. If AI use is the accepted medical standard in a given context, it could be assumed that patients consent implicitly, similar to consenting to technology use when entering a hospital (for example, to undergo a CT scan). Specific consent is only required (to respect patient autonomy) if AI is making decisions independently, if its use deviates from standard use, or if it is *not* used in a context in which the accepted medical standard calls for its use. Explicit consent would also be needed if data are shared outside of the hospital (to comply with data protection legislation).

Hospital authorities and regulators must ensure that doctors and other healthcare professionals are aware of, implement, and comply with the correct standards regarding consent to the use of AI. An insistence on always informing patients about all AI use is disproportionate and could waste the patient's and the professional's time unnecessarily, but in certain circumstances (as described above), its use must be disclosed. Ultimately, AI will become an accepted and routine part of clinical care, but we are not there yet; for now, care must be taken to ensure that patients are given information about AI when required.

Acknowledgments

Authors' contributions: DS wrote the original draft, and all other authors revised it and contributed substantially to the final analysis. BE is the grant holder.

Financial disclosure

The EXPLaiN project was funded by the Swiss National Science Foundation, grant number 407740_187263/1.

Potential competing interests

All authors have completed and submitted the International Committee of Medical Journal Editors form for disclosure of potential conflicts of interest. No potential conflict of interest related to the content of this manuscript was disclosed.

References

- Arbelaez Ossa L, Lorenzini G, Milford SR, Shaw D, Elger BS, Rost M. Integrating ethics in AI development: a qualitative study. *BMC Med Ethics*. 2024 Jan;25(1):10. <http://dx.doi.org/10.1186/s12910-023-01000-0>.
- Lorenzini G, Arbelaez Ossa L, Milford S, Elger BS, Shaw DM, De Clercq E. The "Magical Theory" of AI in Medicine: Thematic Narrative Analysis. *JMIR AI*. 2024 Aug;3:e49795. <http://dx.doi.org/10.2196/49795>.
- Thouvenin F, Bernice Elger, Shaw D, Lorenzini G, Arbelaez Ossa L, Mätzler S; Thouvenin F, Elger B, Shaw D/ Aufklärung beim Einsatz von KI-Systemen in der medizinischen Behandlung: ein Vorschlag auf rechtlicher und rechtstatsächlicher Grundlage. *Jusletter*. 2024;(Januar): <http://dx.doi.org/10.38023/786b14a4-8ce8-452d-a15e-8bd557b0cbcd>.
- Stark L. Medicine's Lessons for AI Regulation. *N Engl J Med*. 2023 Dec;389(24):2213–5. <http://dx.doi.org/10.1056/NEJMp2309872>.
- Ng AY, Oberije CJ, Ambrózy É, Szabó E, Serfőző O, Karpati E, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. *Nat Med*. 2023 Dec;29(12):3044–9. <http://dx.doi.org/10.1038/s41591-023-02625-9>.
- Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial Intelligence in Surgery: promises and Perils. *Ann Surg*. 2018 Jul;268(1):70–6. <http://dx.doi.org/10.1097/SLA.0000000000002693>.
- Gandhi TK, Classen D, Sinsky CA, Rhew DC, Vande Garde N, Roberts A, et al. How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open*. 2023 Aug;6(3):ooad079. <http://dx.doi.org/10.1093/jamiaopen/ooad079>.
- Song B, Zhou M, Zhu J. Necessity and Importance of Developing AI in Anesthesia from the Perspective of Clinical Safety and Information Security. *Med Sci Monit*. 2023 Feb;29:e938835. <http://dx.doi.org/10.12659/MSM.938835>.
- Brown G, Conway S, Ahmad M, Adegbe D, Patel N, Myneni V, et al. Role of artificial intelligence in defibrillators: a narrative review. *Open Heart*. 2022 Jul;9(2):e001976. <http://dx.doi.org/10.1136/openhrt-2022-001976>.
- Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet* 2020 VOLUME 395, ISSUE 10236, P1579-1586, [https://doi.org/http://dx.doi.org/10.1016/S0140-6736\(20\)30226-9](https://doi.org/http://dx.doi.org/10.1016/S0140-6736(20)30226-9).
- Wang F, Preininger A. AI in Health: State of the Art, Challenges, and Future Directions. *Yearb Med Inform* 2019; 28(01): 016-026. DOI: <http://dx.doi.org/10.1055/s-0039-1677908>.
- Spitale G, Schneider G, Germani F, Biller-Andorno N. Exploring the role of AI in classifying, analyzing, and generating case reports on assisted suicide cases: feasibility and ethical implications. *Front. Artif. Intell.*, 14 December 2023. DOI= <http://dx.doi.org/10.3389/frai.2023.1328865>.
- Rezk E, Eltorki M, El-Dakhkhni W. Improving Skin Color Diversity in Cancer Detection: Deep Learning Approach. *JMIR Dermatol*. 2022 Aug;5(3):e39143. <http://dx.doi.org/10.2196/39143>.
- Nuffield Council on Bioethics. Artificial intelligence (AI) in healthcare and research. 2018. <https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research>
- Walker HL, Ghani S, Kuemmerli C, Nebiker CA, Müller BP, Raptis DA, et al. Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument. *J Med Internet Res*. 2023 Jun;25:e47479. <http://dx.doi.org/10.2196/47479>.
- Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. Oxford: Oxford University Press; 1979.
- Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making. *Bioethics*. 2023 Jun;37(5):424–9. <http://dx.doi.org/10.1111/bioe.13158>.
- Coeckelenbergh S, Boelefahr S, Alexander B, Perrin L, Rinehart J, Joosten A, et al. Closed-loop anesthesia: foundations and applications in contemporary perioperative medicine. *J Clin Monit Comput*. 2024 Apr;38(2):487–504. <http://dx.doi.org/10.1007/s10877-023-01111-4>. ; Epub ahead of print.
- Goudra B, Singh PM. Failure of Sedasys: Destiny or Poor Design? *Anesth Analg*. 2017 Feb;124(2):686–8. <http://dx.doi.org/10.1213/ANE.0000000000001643>.
- Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019 Sep;394(10201):861–7. [http://dx.doi.org/10.1016/S0140-6736\(19\)31721-0](http://dx.doi.org/10.1016/S0140-6736(19)31721-0).
- Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022 Apr;4(4):e214–5. [http://dx.doi.org/10.1016/S2589-7500\(22\)00029-2](http://dx.doi.org/10.1016/S2589-7500(22)00029-2).
- Arbelaez Ossa L, Starke G, Lorenzini G, Vogt JE, Shaw DM, Elger BS. Re-focusing explainability in medicine. *Digit Health*. 2022 Feb;8:20552076221074488. <http://dx.doi.org/10.1177/20552076221074488>.
- Pagallo U, O'Sullivan S, Nevejas N, Holzinger A, Friebe M, Jeanquartier F, et al. The underuse of AI in the health sector: opportunity costs, success stories, risks and recommendations. *Health Technol (Berl)*. 2024;14(1):1–14. <http://dx.doi.org/10.1007/s12553-023-00806-7>.