

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE DOCTOR-PATIENT RELATIONSHIP

ESSENTIAL ELEMENTS



Report commissioned by the
Steering Committee for Human Rights
in the fields of Biomedicine and Health (CDBIO)

Author: Brent Mittelstadt

***THE IMPACT OF ARTIFICIAL INTELLIGENCE
ON THE DOCTOR-PATIENT RELATIONSHIP***

By Brent Mittelstadt, Senior Research Fellow and Director of Research at the Oxford Internet Institute, University of Oxford, United Kingdom

All requests concerning the reproduction or translation of all or part of this document should be addressed to the Directorate of Communication (F-67075 Strasbourg Cedex).

All other correspondence concerning this document should be addressed to the Directorate General of Human Rights and Rule of Law.

© Council of Europe, December 2021

ESSENTIAL ELEMENTS

1. In response to a call by the Committee on Bioethics (DH-BIO)¹ to work on trust, safety, and transparency, this report investigates the known and potential impacts of AI systems on the doctor-patient relationship. This impact is framed by the human rights principles referred to in the European Convention on Human Rights and Biomedicine of 1997, otherwise known as the “Oviedo Convention,” and its subsequent amendments.
2. The deployment of AI in clinical care remains nascent. Clinical efficacy has been established for relatively few systems when compared to the significant research activity in healthcare applications of AI. Research, development, and pilot testing often do not translate into proven clinical efficacy, commercialization, or widespread deployment. The generalization of performance from trials to clinical practice generally remains unproven.
3. A defining characteristic of medicine is the ‘healing relationship’ between clinicians and patients. This relationship is augmented by the introduction of AI. However, the role of the patient, the factors that lead people to seek medical attention, and the patient’s vulnerability are not changed by the introduction of AI as a mediator or provider of medical care. Rather, what changes is the means of care delivery, how it can be provided, and by whom. The shift of expertise and care responsibilities to AI systems can be disruptive in many ways.
4. The potential human rights impact of AI on the doctor-patient relationship can be categorised according to six themes: (1) Inequality in access to high quality healthcare; (2) Transparency to health professionals and patients; (3) Risk of social bias in AI systems; (4) Dilution of the patient’s account of well-being; (5) Risk of automation bias, de-skilling, and displaced liability; and (6) Impact on the right to privacy.
5. Concerning (1), as an emerging technology the deployment of AI systems will not be immediate or universal across all member states or healthcare systems. Deployment across institutions and regions will inevitably be inconsistent in terms of scale, speed, and prioritisation.
6. The impact of AI on clinical care and the doctor-patient relationship remains uncertain and will certainly vary by application and use case. AI systems may prove to be more efficient than human care, but also provide lower quality care featuring fewer face-to-face interactions.
7. The inconsistent rollout of AI systems with uncertain impacts on access and care quality poses a risk of creating new health inequalities in member states.

¹ Committee replaced by the Steering Committee for Human rights in the fields of Biomedicine and Health (CDBIO).

8. Article 4 of the Oviedo Convention addresses care provided by healthcare professionals bound by professional standards. It remains unclear whether developers, manufacturers, and service providers for AI systems will be bound by the same professional standards.
9. Careful consideration must be given to the role played by healthcare professions bound by professional standards when incorporating AI systems that interact directly with patients.
10. Concerning (2), transparency and informed consent are key values in the AI-mediated doctor-patient relationship. The complexity of AI raises a question: how should AI systems explain themselves, or be explained, to doctors and patients? This question has many possible meanings: (i) How does an AI system or model function? How was a specific output produced by an AI system? (ii) How was an AI system designed and tested? How is it governed? (iii) What information is required to investigate the behaviour of AI systems? Answers to each of these questions may be necessary to achieve informed consent in AI-mediated care.
11. In cases where AI systems provide some form of clinical expertise, for example by recommending a particular diagnosis or interpreting scans, this requirement to explain one's decision-making would seemingly be transferred from doctor to AI system, or at least to manufacturer of AI system. The difficulty of explaining how AI systems turn inputs into outputs poses a fundamental challenge for informed consent. Aside from the patient's capacity to understand the functionality of AI systems, in many cases patients simply do not have sufficient levels awareness to make free and informed consent possible. AI systems use unprecedented volumes of data to make their decisions, and interpret these data using complex statistical techniques, both of which increase the difficulty and effort required to remain aware of the full scope of data processing and clinical analysis informing one's diagnosis and treatment.
12. AI systems interacting directly with patients should self-identify as an artificial system. Whether the usage of AI systems in care settings should always be disclosed to patients by clinicians and healthcare institutions is a more difficult question.
13. Concerning (3), AI systems are widely recognised as suffering from bias in their inputs, processing, and outputs. Biased and unfair decision-making often occurs not for technical or regulatory reasons, but rather reflects underlying social biases and inequalities. For example, samples in clinical trials and health studies have historically been biased towards white male subjects meaning results are less likely to apply to women and people of colour.
14. Social biases in AI systems can lead to unequal distribution of outcomes across patient populations and protected demographic groups. Western societies have long been marked by significant social inequality. These historical and contemporary trends influence the training of future systems. Without

intervention, these patterns in access to healthcare opportunities and resources will be learned and reinforced by AI systems.

15. Detecting biases in AI systems is not straightforward. Biased decision-making rules can be hidden in 'black box' models. Simply anonymising health data may not be an adequate solution to mitigate biases due to the influence of historical inequality and the existence of strong proxies for protected attributes (e.g., post code as a proxy for ethnicity). The various challenges of social bias, discrimination, and inequality suggest health professionals and institutions face a difficult task in ensuring their usage of AI systems does not further existing inequalities and create new forms of discrimination.
16. Concerning (4), the development of trust in a doctor-patient relationship may be inhibited by technological mediation. As a mediator placed between the doctor and patient, AI systems can inhibit tacit understanding of the patient's health and well-being and encourage both clinician and patient to discuss health solely in measurable quantities or machine interpretable terms.
17. Concerning (5), to ensure patient safety and replace the protection offered by human clinical expertise, robust testing and validation standards should be an essential pre-deployment requirement for AI systems in clinical care contexts. Evidence of clinical efficacy does not yet exist for many AI applications in healthcare, which has justifiably proven a barrier to widespread deployment.
18. Concerning (6), AI poses several unique challenges to the human right to privacy and complementary data protection regulations. These rights seek to provide individuals with greater transparency and control over automated forms of data processing. They will undoubtedly provide valuable protection for patients across a variety of use cases of medical AI.
19. The Oviedo Convention sets out a specific application of the right to privacy (Article 8 ECHR) which recognises the particularly sensitive nature of personal health information and sets out a duty of confidentiality for health care professionals.
20. Ethical standards need to be developed around transparency, bias, confidentiality, and clinical efficacy to protect patient interests in informed consent, equality, privacy, and safety. Such standards could serve as the basis for deployments of AI in healthcare that help rather than hinder the trusting relationship between doctors and patients.
21. Where AI can be observed to have a clear impact on rights and protections set out in the Oviedo Convention, it is appropriate for the Council of Europe to introduce binding recommendations and requirements for signatories concerning how AI is deployed and governed. Recommendations should focus on a higher positive standard of care with regards to the doctor-patient relationship to ensure it is not unduly disrupted by the introduction of AI in care settings.

22. The Council of Europe could set standards for what and how information about the recommendation of an AI system concerning a patient's diagnosis and treatment should be communicated to the patient. These standards should likewise address the doctor's role in explaining AI recommendations to patients and how AI systems can be designed to support the doctor in this role.
23. The capacity of AI to replace or augment human clinical expertise utilising highly complex analytics and unprecedented volumes and varieties of data suggests its impact on the doctor-patient relationship may be unprecedented.
24. The degree to which AI systems inhibit 'good' medical practice hinges upon the model of service. If AI is used solely to complement the expertise of health professionals bound by the fiduciary obligations of the doctor-patient relationship, the impact of AI on the trustworthiness and human quality of clinical encounters may prove to be minimal. At the same time, if AI is used to heavily augment or replace human clinical expertise, its impact on the caring relationship is more difficult to predict. It is entirely possible that new, broadly accepted norms for 'good' care will emerge through greater reliance on AI systems, with clinicians spending more time face-to-face with patients and relying heavily on automated recommendations. The impact of AI on the doctor-patient relationship nonetheless remains highly uncertain. We are unlikely to see a radical reconfiguration of care in the next five years in the sense of human expertise being replaced outright by artificial intelligence.
25. A radical reconfiguration of the doctor-patient relationship of the type imagined by some commentators, in which artificial systems diagnose and treat patients directly with minimal interference from human clinicians, continues to seem far in the distance.
26. Going forward, the ideal model of clinical care and AI deployment in healthcare is one that utilises the best aspects of human clinical expertise and AI diagnostics.
27. The doctor-patient relationship is a keystone of 'good' medical practice, and yet it is seemingly being transformed into a doctor-patient-AI relationship. The challenge facing AI providers, regulators, and policymakers is to set robust standards and requirements for this new type of 'healing relationship' to ensure patients' interests and the moral integrity of medicine as a profession are not fundamentally damaged by the introduction of AI.