Creative Computing Institute
University of the Arts London

Ethical Dilemmas of Artificial Intelligence in Healthcare: Balancing Progress and Principles

Shaan Rehsi
January 27 2024
Computational Ethics
Year 3

Contents

Introduction

As artificial intelligence (AI) continues to develop, we can expect it will play an increasingly significant role within the healthcare sector. There is already an understanding of the uses of AI, for example, the American Hospital Association (AHA) highlight that AI can improve clinical decision-making, and diagnostics and lead to improved patient safety by optimising health outcomes (American Hospital Association, 2023). Despite this, the U.S. Government Accountability Office (GAO) have indicated challenges such as data access, bias, and privacy concerns, emphasising the need for careful integration of these technologies (The U.S. Government Accountability Office, 2021).

This essay explores the evolving role of AI in healthcare, looking into the benefits it brings to the medical field and addressing three main ethical concerns: the trust between patients and doctors, reflecting the principle of beneficence; data security and privacy, which ties into the respect for patient autonomy; and accountability and transparency in AI use, which are crucial for ensuring justice and non-maleficence. The essay concludes by discussing the frameworks that need to be put into place to manage ethical concerns. As AI's footprint in healthcare deepens, it will be crucial to address these concerns head-on, ensuring that this technological leap forward does not compromise the core values of medical practice.

Positive Cases of Artificial Intelligence in Healthcare

The use of AI in healthcare, particularly in cancer diagnosis and treatment, has been transformative. A notable example is an AI-assisted endoscopic system developed in Japan, designed to aid in early cancer detection (Mori et al., 2021). In one of the first trials of AI-assisted endoscopy in a clinical setting, the system demonstrated a high degree of effectiveness in identifying colorectal tumours during colonoscopies. The AI system was tested as part of a clinical study of 700 patients at Showa University. All patients underwent a colonoscopy, they found that the AI was able to detect pre-cancerous lesions with an accuracy greater than 90% (Eureka Alert, 2023).

In a separate study involving AI and dermatologists, an AI system outperformed most dermatologists in detecting skin cancer. The AI, trained with over 100,000 images, achieved a 95% accuracy rate in identifying skin cancers, compared to 87% for human dermatologists (Pham et al., 2021). The study, which included dermatologists of varying experience levels, demonstrated the AI's higher sensitivity and fewer misdiagnoses (Presse, A., 2018.). This suggests AI's potential as a tool for quicker, more reliable skin cancer diagnoses, although experts emphasise that it cannot replace thorough clinical examinations.

Researchers at Imperial College London are currently leading a study to investigate AI's role in enhancing breast cancer screening and diagnosis. Collaborating with DeepMind Health, Google's

AI health research team, and a consortium of experts, the project goal is to leverage AI and machine learning to enhance the accuracy of breast cancer detection on mammograms. If successful, this collaboration could lead to more efficient and effective cancer screening services, potentially reducing late-stage diagnoses, and improving patient outcomes. While the project is in its initial stages, the involvement of leading AI experts and healthcare professionals holds promise for advancing breast cancer diagnosis and care through technology (O'Hare, 2017).

Ethical Concerns: Trust, Privacy and Accountability

In contrast to these advancements, many patients have concerns with the integration of AI systems in healthcare, specifically around the patient-physician relationship highlighting trust, and the human element in care. Historically, the medical field has depended on the relationship between patients and their physicians, which is built on trust, empathy, and direct human-to-human interaction (Kerasidou, 2020). However, the rise of AI tools in diagnosis, treatment recommendations, and patient monitoring could potentially disrupt this dynamic. There is potential that a patient may trust decisions made by AI over their physician due to an AI only analysing the given information and not being prone to human infallibility (Dalton-Brown, 2020).

While AI can aid a physician's capabilities, there is a risk that an over-reliance on technology might lead to a perception of a less personalised approach to care. In addition, patients may resent the integration of AI; the idea of their healthcare experience becoming more transactional, where

patients might feel less engaged or understood by their healthcare providers, which can erode trust (Holtz, Nelson, & Poropatich, 2023).

The over-reliance on AI systems used in healthcare directly challenges the ethical principle of beneficence. In the context of healthcare, beneficence guides medics to not only act as they believe is in the best interest of the patients through effective treatments but also ensure that patients are fully informed and involved in decisions about their healthcare (Premela, 2023). When doctors overuse AI systems this goes directly against beneficence as patients are unable to fully understand or contribute to decisions regarding their health, thus impinging upon their autonomy and potentially leading to decisions that may not align with their values or best interests. This situation calls for a careful balance in AI usage, upholding beneficence, trust, and transparency.

Additionally, AI falls short of understanding and expressing human empathy, a crucial element in healthcare (Kerasidou, 2020). Empathy often arises from shared experiences, something AI cannot replicate (Dalton-Brown, 2020). The European Parliament Committee highlights the irreplaceable value of human interaction in healthcare, warning against the dehumanisation of care through the replacement of humans with robots (European Parliament, Committee on Legal Affairs, 2018). While AI contributes to accuracy and efficiency, its limitations in empathy challenge the principle of beneficence, risking harm to patient well-being. To uphold beneficence, it is essential to strike a balance where AI enhances but does not replace human

elements, ensuring patients receive empathetic and personalised care in their healthcare journey (Dalton-Brown, 2020).

Trust in the patient-physician relationship is foundational, yet this trust extends beyond personal interactions to the security and confidentiality of medical data. This area is equally crucial and demands attention as it directly impacts patient autonomy. Patient autonomy includes the right to control one's personal health information and can be split up into two categories. Firstly, authenticity, in the context of patient autonomy, entails that patients' decisions and consents are based on their genuine beliefs and values, free from external influences of AI systems (Prunkl, 2022). The second concept is agency which in the context of autonomy emphasises the patient's ability to act based on their values and decisions. This is crucial in healthcare and will need to be navigated when utilising AI to ensure patients' rights are respected (Gerke, et al., 2020). However, AI's opaque use in healthcare may limit agency by processing patient data in ways that are not fully transparent or understood by the patients themselves (von Eschenbach, 2021). This lack of transparency can impede patients' ability to make decisions that are truly their own, thereby the authenticity aspect of autonomy.

Concerns about patient privacy in AI-driven healthcare directly impact patient autonomy. A 2019 Pew survey where 81% of U.S. respondents expressed fears about data collection risks (Null et al., 2019). AI-driven predictive models used in diagnostics and treatment planning require access to detailed patient records, which could include everything from medical history to genetic information (Topol, 2019). If AI systems experienced data breaches or unauthorised access, in

addition to confidential health information being at risk, the misuse of this data could be detrimental, for example, discriminatory practices in insurance or employment (Morley, 2023).

Additionally, a survey surrounding the research of trust in AI surveyed 408 UK hospital patients on AI awareness and data sharing for AI research. Initially, most had limited AI understanding. After explanations, 78% agreed to share health data with the NHS, 66% with universities, but only 26% with commercial entities, citing misuse concerns (Aggarwal et al., 2021). The study highlights the importance of patient awareness in AI healthcare applications. As patients become more informed about AI's role, they show greater autonomy, including a willingness to share data with trusted entities like the NHS. Enhanced transparency, education about AI's use, and trust-building are key to supporting patient autonomy. This approach respects patients' rights to control their health information, crucial in patient-centred healthcare as AI integration grows.

An example of this in recent years was the data transfer of NHS patient records to DeepMind, Google's AI company. DeepMind was given access to the records of 1.6 million patients to develop a healthcare app called Streams. There has been controversy surrounding this agreement as the Information Commissioner's Office, the UK's data watchdog found that DeepMind did not gain the proper patient consent to use their data (Revell, 2017). The handling of sensitive patient data without adequate transparency raises significant privacy concerns. In the example of DeepMind, the lack of informed consent meant that patients were unaware and unable to make an informed decision about their data, which is a fundamental requirement in healthcare ethics (Rigby, 2019). This situation underscores the need for clear regulations around data sharing

between healthcare providers and AI entities. When AI processes patient data without explicit consent, it challenges the authenticity aspect of autonomy, as patients are not making fully informed decisions based on their values and beliefs.

Navigating the complexities of patient data privacy highlights additional issues with the integration of AI in healthcare such as AI-assisted decision-making. It is crucial that these AI systems, entrusted with this data and important decisions, operate within a framework of accountability transparency and justice. AI systems must be held accountable for their decisions and any harm caused, adhering to healthcare ethics and regulations (von Eschenbach, 2021).

When decisions are influenced or made entirely by AI systems, it becomes essential that patients understand the basis of these decisions. This understanding is exacerbated by the lack of transparency in AI decision-making processes, often referred to as the 'black box' issue. This issue arises because the internal workings and decision-making logic of AI systems are not always visible or understandable to users, including healthcare providers. This opacity makes it difficult to identify and correct errors or biases in AI models (Wadden, 2022). For example, an AI trained on MRI scans from specific machines has shown a decreased accuracy in diagnoses when applied to data from different machines (Martin-Isla, et al., 2020). Such discrepancies were also observed in pneumonia diagnosis AI systems across different hospitals (Zech et al., 2018) and in DeepMind's retinal disease diagnosis model (De Fauw et al., 2018), where the error rate increased when applied to new datasets.

Ethically, patients must be afforded a clear understanding of how decisions regarding their health are determined. The opacity often associated with AI systems stands in violation of this right. Such transparency is the foundation of justice, guaranteeing that patients are accorded equitable treatment and dignity. This principle asserts that clarity in healthcare decision-making is not only a matter of ethical obligation but also a requisite for ensuring the just and respectful treatment of patients (von Eschenbach, 2021).

The principle of non-maleficence or “do no harm", is challenged by AI systems. This principle dictates that actions taken by healthcare professionals and systems, should not cause harm or injury to patients (Premela, 2023). In practice and medical decision-making, it means that the welfare of the patient is a primary consideration, and any treatment or intervention should not adversely affect them (Summers & Morrison, 2009). This can be illustrated by IBM Watson, a system designed to deliver lung cancer treatment recommendations. These recommendations were sometimes based on hypothetical treatment scenarios rather than real-world data, leading to potential patient harm and questioning the system’s reliability (Ross & Swetlitz, 2017). The system often struggled with complex data, resulting in misaligned medical recommendations (Strickland, 2019), conflicting with the non-maleficence principle, and raising questions of who should be responsible when AI in healthcare makes an error, whether it is the developers, the healthcare providers, or the AI system itself. This emphasises the need for AI systems to continuously learn from diverse clinical data for accurate patient care (Topol, 2019) and highlights the challenge of assigning responsibility when AI advice is suboptimal. Ensuring human oversight in AI decision-making is crucial for ethical medical practice and upholding justice.

AI-guided systems used in healthcare may also contribute to non-maleficence and may cause issues due to errors in the data they are trained with, leading to misdiagnoses or inappropriate treatments (Challen et al., 2019; Ellahham et al., 2020). Issues like false positives can arise from poor data quality, such as inaccuracies in ultrasound scans influenced by operator experience and patient cooperation (Pinto et al., 2013). These factors contribute to the challenge of assigning accountability for AI-driven errors in healthcare.

Conclusion

To conclude, this essay has highlighted the integration of AI in healthcare presents both remarkable opportunities and significant ethical challenges. AI's potential to revolutionise medical diagnostics and treatment in cancer detection and patient care was emphasised. On the other hand, many ethical concerns have arisen, including issues surrounding data security, patient privacy, and the evolving patient-physician relationship. These aspects simultaneously validate the American Hospital Association's (AHA) support for AI in healthcare and align with the U.S. Government Accountability Office's (GAO) cautions regarding data protection. These challenges, if not addressed with thorough and robust frameworks, could compromise the fundamental methods and values of medical practice.

The integration of AI within the healthcare sector should harness AI's capabilities to enhance patient outcomes while upholding the highest ethical standards. As AI is integrated more into the

medical field, healthcare providers, policymakers, and AI developers must work collaboratively to develop transparent data governance policies and rigorous oversight mechanisms. They must navigate these complexities, striking a balance between technological innovation and the core values of medical practice. By doing so, AI can become a true ally in healthcare, transforming the medical sector while respecting and protecting the fundamental rights and dignity of patients.

Word Count: 2276

References

- Aggarwal, R., Farag, S., Martin, G., Ashrafian, H. and Darzi, A., 2021. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. *Journal of Medical Internet Research*, 23(8), p.e26162.
- American Hospital Association (AHA), 2023. How AI Is Improving Diagnostics, Decision-Making and Care. AHA Center for Health Innovation Market Scan, [online] 9 May. Available at: <https://www.aha.org/aha-center-health-innovation-market-scan/2023-05-09-how-ai-improving-diagnostics-decision-making-and-care> [Accessed 27 December 2023].
- Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T. and Tsaneva-Atanasova, K., 2019. Artificial intelligence, bias and clinical safety. BMJ quality & safety.
- Dalton-Brown, S., 2020. The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics*, 29(1), pp.115-121.

- De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D. and Van Den Driessche, G., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine, 24(9), pp.1342-1350.
- Ellahham, S., 2020. Artificial intelligence: the future for diabetes care. The American journal of medicine, 133(8), pp.895-900.
- Eureka Alert (2023) Artificial Intelligence: Is this the future of early cancer detection?, EurekAlert! Available at: https://www.eurekalert.org/news-releases/605415 (Accessed: 27 January 2024).
- Gerke, S., Minssen, T. and Cohen, G. (2020) 'Ethical and legal challenges of artificial intelligence-driven healthcare', Artificial Intelligence in Healthcare, pp. 295–336. doi:10.1016/b978-0-12-818438-7.00012-5.
- Holtz, B., Nelson, V., Poropatich, R.K., 2023. Artificial Intelligence in Health: Enhancing a Return to Patient-Centered Communuication. *Telemedicine and e-Health*, 29(6), pp.795-797.
- Kerasidou, A., 2020. Artificial intelligence and the ongoing need for empathy, compassion, and trust in healthcare. *Bulletin of the World Health Organization*, 98(4), p.245.
- Martin-Isla, C., Campello, V.M., Izquierdo, C., Raisi-Estabragh, Z., Baeßler, B., Petersen, S.E. and Lekadir, K., 2020. Image-based cardiac diagnosis with machine learning: a review. Frontiers in cardiovascular medicine, 7, p.1.

- Mori, Y. et al. (2021) ‘Artificial Intelligence-assisted colonic endocytoscopy for cancer recognition: A multicenter study’, Endoscopy International Open, 09(07). doi:10.1055/a-1475-3624.
- Morley, J., 2023. Thinking Critically about AI in Healthcare. Oxford Internet Institute, University of Oxford, May.
- Null, E., Oribhabor, I. and Escoto, W., 2021. Data minimisation is key to protecting and reducing harm. [online] Access Now. Available at: <https://www.accessnow.org/cms/assets/uploads/2021/05/Data-Minimization-Report.pdf> [Accessed 14 December 2023].
- O’Hare, R. (2017) Research collaboration aims to improve breast cancer diagnosis using AI: Imperial News: Imperial College London, Imperial News. Available at: https://www.imperial.ac.uk/news/183293/research-collaboration-aims-improve-breast-cancer/ (Accessed: 27 January 2024).
- Pham, T.C., Luong, C.M., Hoang, V.D., et al., 2021. AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function. *Scientific Reports*, 11, 17485.
- Pinto, A. et al. (2013) ‘Sources of error in emergency ultrasonography’, Critical Ultrasound Journal, 5(S1). doi:10.1186/2036-7902-5-s1-s1.
- Presse, A., 2018. Skin cancer: Computer learns to detect skin cancer more accurately than a doctor. [online] Available at: <https://www.theguardian.com/society/2018/may/29/skin-cancer-computer-learns-to-detect-skin-cancer-more-accurately-than-a-doctor> [Accessed 20 January 2024].

- Premela, 2023. Medical Ethics: Beneficence. *The Medic Portal*, [online] 12 Dec. Available at: <www.themedicportal.com/application-guide/medical-school-interview/medical-ethics/medical-ethics-beneficence/> [Accessed 14 December 2023].
- Premela, 2023. Medical Ethics: Non-Maleficence. *The Medic Portal*, [online] 12 Dec. Available at: <www.themedicportal.com/application-guide/medical-school-interview/medical-ethics/medical-ethics-non-maleficence/> [Accessed 14 December 2023].
- Prunkl, C., 2022. Human autonomy in the age of artificial intelligence. *Nature Machine Intelligence*, 4(2), pp.99-101.
- Revell, T., 2017. Google DeepMind's NHS data deal 'failed to comply' with law. *New Scientist*. [online] Available at: <https://www.newscientist.com/article/2139395-google-deepminds-nhs-data-deal-failed-to-comply-with-law/> [Accessed 20 December 2023].
- Rigby, Michael J., 2019. Ethical Dimensions of Using Artificial Intelligence in Health Care. *Journal of Ethics | American Medical Association*, [online] American Medical Association, 1 Feb. Available at: <journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02> [Accessed 30 December 2023].
- Ross, C. and Swetlitz, I., 2017. IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *Stat*.
- Strickland, E., 2019. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4), pp.24-31.
- Summers, J. and Morrison, E., 2009. Principles of healthcare ethics. *Health Care Ethics*. 2nd ed. Sudbury: Jones and Bartlett Publishers, pp.41-58.

- The EU Legal Affairs Committee, 2018. Report with recommendations to the Commission on Civil Law Rules on Robotics. [online] Available at: <http://europa.eu/rapid/press-release_IP-18-3362_en.htm> [Accessed 12 January 2023].
- Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), pp.44-56.
- U.S. Government Accountability Office, 2021. Artificial Intelligence in Health Care: Benefits and Challenges of Technologies to Augment Patient Care, GAO-21-7SP. [online] Available at: <https://www.gao.gov/products/gao-21-7sp> [Accessed 27 December 2023].
- von Eschenbach, W.J., 2021. Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), pp.1607-1622.
- Wadden, J.J., 2022. Defining the undefinable: the black box problem in healthcare artificial intelligence. *Journal of Medical Ethics*, 48(10), pp.764-768.
- Zech, J.R. et al. (2018) 'Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study', PLOS Medicine, 15(11). doi:10.1371/journal.pmed.1002683.

AI Log

Instance	AI Tool Used	Purpose	Section of Essay

1	Chat GPT 3.5	Condensing words for clarity	Introduction
2	Chat GPT 3.5	Cohesion to introducing each section of my essay	Pages 7 & 9
3	Chat GPT 3.5	Summarising and structuring	Conclusion