Untangling The Ethical Labyrinth Of Artificial Intelligence in Medicine

The burgeoning field of artificial intelligence (AI) – defined as technologies capable of imitating human intelligence – has presented an array of transformative possibilities across various fields, with medicine being no exception (Merriam-Webster, 2023; Laylo, 2023). Indeed, the rapid integration of AI technologies such as large language models and machine learning algorithms has immense potential to revolutionise diagnostics and treatment planning (van de Sande et al, 2022; Topol, 2019). However, the incorporation of artificial intelligence in medicine introduces a unique labyrinth of complex ethical dilemmas. This essay will thus endeavour to dissect the key ethical considerations that emerge from the implementation of AI in medicine through the lens of the relevant ethical theories. Particularly, this essay will highlight the central tension between the utilitarian drive to optimise patient outcomes and the deontological imperative to adhere to key moral duties and principles. Whilst the normative ethical theories of utilitarianism and deontology are heterogonous philosophies with considerable internal variance, for the purpose of this essay they will be considered in terms of their core doctrines.

As a consequentialist philosophy, utilitarianism is ultimately concerned with maximising 'utility' – characterised by preeminent utilitarian thinker Jeremy Bentham as 'benefit, advantage, pleasure, good, or happiness' (Bentham, 1789). Accordingly, a utilitarian approach to determining the ethicality of implementing AI in medicine would ultimately comprise of an estimation of its net utility – essentially weighing its benefits against its harms. Through such a lens, the ethical case for AI-driven medicine is compelling given the widely accepted capability of AI tools to positively revolutionise patient outcomes across all medical fields (Pereira et al, 2023). For instance, the deep complexity of learning models enables the identification of patterns in clinical data that would otherwise be uninterpretable by human clinicians and can be used to guide diagnostic and/or treatment decisions to improve quality of care (Briganti and Le Moine, 2020). Alongside enhancing patient care through earlier detection and diagnosis, AI-tools have the power to improve workflow (thereby reducing medical errors) and considerably decrease morbidity and mortality (Mintz and Brodie, 2019). Although difficult to quantify, AI implementation thus appears inevitable to generate vast utility. Therefore, since utilitarianism is concerned with net utility, the integration of AI into medicine would be considered ethically virtuous notwithstanding any potential detrimental consequences (e.g. doctors losing their jobs) – so long as the disutility of any negative effects does not outweigh the immense utility generated.

Contrariwise, deontological ethics foregrounds adherence to moral duties and principles (Alexander and Moore, 2020). Accordingly, a deontological approach would examine the ethicality of AI integration in medicine by assessing the extent to which AI-driven solutions adhere to relevant moral duties and principles. Through this lens, the ethical implications of AI in medicine are somewhat murkier, with AI-tools potentially threatening the core principles of autonomy, accountability, justice, and the patient-doctor relationship. The specific implications of AI-driven medicine for these key principles of medical ethics will thus be explored to gain insight into the deontological perspective.

Autonomy, one of the four pillars of medical ethics, emphasises the right of individuals to make informed decisions about their healthcare (Beauchamp and Childress, 2009). It has been suggested that AI-driven healthcare could increase patient's autonomy by enabling patients to more directly engage in their healthcare through accessible apps or wearable technologies (Briganti and Le Moine, 2020). However, increased reliance on AI may diminish patients' agency over healthcare decisions, shifting decision-making power to

AI systems and thus infringing autonomy (Mittelstadt et al, 2016). Additionally, a necessary tenet of autonomy is patients being well-informed (i.e. understanding the reasoning behind a particular diagnosis or treatment recommendation), which is threatened by AI-driven medicine due to the complexity and opacity of AI algorithms (Char et al, 2018). Indeed, many AI-driven algorithms operate as 'black boxes' with their internal workings being inaccessible to human scrutiny, making it difficult for doctors and patients alike to understand or challenge algorithmic decisions. As an illustration of this phenomenon, Castelvecchi (2016) discusses the implications of a deep-learning neural network trained using old mammograms that can identify whether the breast tissue of an apparently healthy woman 'looks' cancerous by implicitly recognising unknown predictive markers, but is unable to explain why. Indeed, it is challenging enough for patients' to elect to undergo a preventative mastectomy due to known genetic risk variants, but this choice would be further convoluted if the patient is unable to know what the risk factor is – notwithstanding whether the algorithm's predictions are generally accurate.

Another issue with the opaque and uninterpretable decision-making process of many AI-algorithms is the consequent absence of accountability. Whether assigned to individual doctors or healthcare systems, accountability is another key principle of ethical healthcare (Bell et al, 2011). Yet, the lack of transparency of AI algorithms complicates any efforts to hold developers, clinicians, or healthcare institutions accountable for AI-driven decisions. For instance, whilst in the traditional model of medical accountability physicians can be held liable in the case of malpractice or negligence, it may become challenging to pinpoint the cause of a mistake and/or appropriately assign responsibility when AI is involved in medical decisions which result in errors or poor outcomes. Indeed, there is already debate over whether liability for AI-related mistakes should lie with the developers of the AI-systems, the providers employing these technologies, or even the AI itself (Maliha et al, 2021).

Justice – another of the tetradic pillars of medical ethics – encompasses the fair and equitable distribution of healthcare resources (Beauchamp and Childress, 2009). AI certainly has the potential to enhance the efficiency/effectiveness of medical care – thus enabling more equitable distribution of limited resources (Morley et al, 2020). For instance, diagnostics driven by AI-algorithms could alleviate the burden on overworked clinicians and facilitate quicker access to treatment for patients in resource-constrained settings (Wartman and Combs, 2018). However, the implementation of AI in medicine also raises concerns about exacerbating existing healthcare disparities. Indeed, unequal access to technologies due to geographic or socioeconomic inequalities could further disadvantage underserved populations (Emanuel and Wachter, 2019). For instance, rural or underserved communities may lack the necessary infrastructure or resources to adopt AI technologies, potentially widening the care quality gap within populations. Moreover, biases embedded in AI algorithms arising from historical data or flawed training datasets could perpetuate discriminatory practices in healthcare (Joyce et al, 2021). Notably, Obermeyer et al (2019) found that an AI-algorithm widely used to assign health risk (and thus guide healthcare decisions) in the US significantly underestimated the medical needs of Black patients as a consequence of using the amount of money spent on patients as a proxy for health needs – failing to account for systematic biases that mean less money is spent on Black patients with the same level of need. This algorithm halved the number of Black patients identified as needing extra care – providing a compelling example of the potential for AI algorithms to perpetuate existing disparities in healthcare access and outcomes.

AI-driven medicine also has significant implications for the patient-doctor relationship. As AI-driven tools become more pervasive, concerns arise about eroding patient-doctor relationships and the associated empathy and human connection (Pesapane et al, 2018). Alongside threatening the shared decision-making process between doctors and

patients, this could risk losing the ‘human touch’ – a historically and culturally central component of healthcare (Drouin and Freeman, 2020).),

Considering the above explorations into the implications of AI-driven medicine for the relevant core principles of medical ethics, the deontological case for opposing the implementation of artificial intelligence into medicine is clear. Indeed, given deontology’s focus on the adherence to key moral duties and obligations notwithstanding the ‘greater good’, the considerable likelihood of AI-driven medicine violating the various key moral principles of medical ethics discussed would preclude the deployment of AI in medicine from being considered ethically virtuous from a deontological perspective.

Having established the utilitarian case for supporting AI-driven medicine (on the basis that its considerable and widespread positive impact on patient outcomes outweighs any detrimental effects) and the deontological case for opposing it (due to its violation of sacrosanct principles of medical ethics), the fundamental tension between these two ethical theories at the heart of the ethical labyrinth of AI healthcare is elucidated. This account does not consider the important nuance that the implementation of AI in medicine is not a unitary interposition and is not all-or-nothing (meaning either of the theories could support the implementation of AI in some capacities and not in others). It also rests on the unprovable assumption that the extent of the utility generated by AI is likely to outweigh the disutility generated, as well as the supposition that the principles discussed are amongst the prima facie ethical imperatives of medical ethics. Nonetheless, this account does offer a useful conceptual characterisation of the overarching conflict between utilitarianism and deontology in the ethical consideration of artificial intelligence in medicine.

By understanding the conflict of these ethical doctrines, the ‘ethical labyrinth’ of AI in medicine can be effectively navigated by attempting to reconcile them. Accordingly, a successful approach to their reconciliation would necessitate the establishment of guidelines, regulatory frameworks, and technological innovations that uphold deontological principles whilst still harnessing the utility generated by AI. Such measures could include improving the interpretability of AI algorithms through the development of explainable AI techniques, establishing a clear framework delineating liability in the case of AI-related errors, legislatively dictating that AI algorithms must be implemented transparently and equitably, and developing shared-decision making models that incorporate both AI-generated recommendations and human input.

In conclusion, the ethical complexities surrounding AI in medicine can largely be elucidated by dissecting the tension between the utilitarian drive to optimise patient outcomes and the deontological imperative to uphold key moral duties and principles. By untangling this ethical labyrinth and identifying particular areas of contention, there may be scope to reconcile these ostensibly antagonistic philosophies to maximise the vast potential benefits of AI in medicine whilst ensuring responsible and ethical integration.

References

Alexander, L. and Moore, M. (2020) *Deontological ethics, Stanford Encyclopedia of Philosophy*. Stanford University. Available at: https://plato.stanford.edu/entries/ethics-deontological/ (Accessed: April 21, 2023).

Beauchamp, T.L. and Childress, J.F. (2009) *Principles of Biomedical Ethics*. New York: Oxford University Press.

Bell, S.K. *et al.* (2011) "Accountability for medical error," *Chest*, 140(2), pp. 519–526. Available at: https://doi.org/10.1378/chest.10-2533.

Bentham, J. (1789) *An introduction to the principles of morals and legislation. printed in the year 1780, and now first published. by Jeremy Bentham* .. London: Printed for T. Payne, and Son.

Briganti, G. and Le Moine, O. (2020) "Artificial Intelligence in medicine: Today and Tomorrow," *Frontiers in Medicine*, 7. Available at: https://doi.org/10.3389/fmed.2020.00027.

Castelvecchi, D. (2016) "Can we open the black box of ai?," *Nature*, 538(7623), pp. 20–23. Available at: https://doi.org/10.1038/538020a.

Char, D.S., Shah, N.H. and Magnus, D. (2018) "Implementing machine learning in health care — addressing ethical challenges," *New England Journal of Medicine*, 378(11), pp. 981–983. Available at: https://doi.org/10.1056/nejmp1714229.

Drouin, S.F. and Freeman, d (2020) *Health care needs ai. it also needs the human touch*, *STAT*. Available at: https://www.statnews.com/2020/01/22/health-care-needs-ai-it-also-needs-human-touch/.

Drouin, S.F. and Freeman, d (2020) *Health care needs ai. it also needs the human touch*, *STAT*. Available at: https://www.statnews.com/2020/01/22/health-care-needs-ai-it-also-needs-human-touch/ (Accessed: April 21, 2023).

Emanuel, E.J. and Wachter, R.M. (2019) "Artificial Intelligence in health care," *JAMA*, 321(23), p. 2281. Available at: https://doi.org/10.1001/jama.2019.4914.

Joyce, K. *et al.* (2021) "Toward a sociology of Artificial Intelligence: A call for research on inequalities and structural change," *Socius: Sociological Research for a Dynamic World*, 7, p. 237802312199958. Available at: https://doi.org/10.1177/2378023121999581.

Laylo, Q (2023) *View of artificial intelligence and its reasons as a result of the Fourth Industrial Revolution, Imras Journal* . Imras. Available at: https://journal.imras.org/index.php/sps/article/view/155/175 (Accessed: April 21, 2023).

Maliha, G. *et al.* (2021) "Artificial Intelligence and liability in Medicine: Balancing Safety and Innovation," *The Milbank Quarterly*, 99(3), pp. 629–647. Available at: https://doi.org/10.1111/1468-0009.12504.

Mintz, Y. and Brodie, R. (2019) "Introduction to artificial intelligence in medicine," *Minimally Invasive Therapy & Allied Technologies*, 28(2), pp. 73–81. Available at: https://doi.org/10.1080/13645706.2019.1575882.

Mittelstadt, B.D. *et al.* (2016) "The Ethics of Algorithms: Mapping the debate," *Big Data & Society*, 3(2), p. 205395171667967. Available at: https://doi.org/10.1177/2053951716679679.

Morley, J. *et al.* (2020) “The ethics of AI in health care: A mapping review,” *Social Science & Medicine*, 260, p. 113172. Available at: https://doi.org/10.1016/j.socscimed.2020.113172.

Obermeyer, Z. *et al.* (2019) “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, 366(6464), pp. 447–453. Available at: https://doi.org/10.1126/science.aax2342.

Pereira, T., Cunha, A. and Oliveira, H.P. (2023) “Special issue on novel applications of Artificial Intelligence in medicine and health,” *Applied Sciences*, 13(2), p. 881. Available at: https://doi.org/10.3390/app13020881.

Pesapane, F. *et al.* (2018) “Artificial Intelligence as a medical device in radiology: Ethical and regulatory issues in Europe and the United States,” *Insights into Imaging*, 9(5), pp. 745–753. Available at: https://doi.org/10.1007/s13244-018-0645-y.

Sande, D.van de *et al.* (2022) *Developing, implementing and governing Artificial Intelligence in medicine: A step-by-step approach to prevent an artificial intelligence winter*, *BMJ Health & Care Informatics*. BMJ Publishing Group Ltd. Available at: https://doi.org/10.1136%2Fbmjhci-2021-100495.

Topol, E. (2019) “High-performance medicine: The convergence of human and Artificial Intelligence,” *Nature Medicine*, 25(1), pp. 44–56. Available at: https://doi.org/10.1038/s41591-018-0300-7.

Wartman, S.A. and Combs, C.D. (2018) “Medical education must move from the information age to the age of Artificial Intelligence,” *Academic Medicine*, 93(8), pp. 1107–1109. Available at: https://doi.org/10.1097/acm.0000000000002044.