Impact of AI and Deep Learning in Medicine and Healthcare

Figgy Smith

20 AUGUST 2025

Introduction

The ever-present myopia for interest and convenience via the mass adoption of AI and Deep Learning will have long term detrimental consequences upon the field of medicine and healthcare, if not tightly regulated with human primacy as the core tenet of these systems. The AI boom has been ever-present and besieging the contemporary discourse; See (Rajpurkar, 2022) At this nascent stage we must make careful consideration of how we use this technology. It ought to be applicable, effective, and ethical especially when matters of public health and care are on the line. We cannot afford a denigration of the field with a multiplicity of issues and threats to the health and well-being of people at large. This simulacrum of intelligence and agency produced by this technology has enthralled many minds. There has been and is a good impact for this field. Although the current path taken shows that despite specific applications within the technical side of healthcare such as data analysis and big data, for example AlphaFold by DeepMind for protein structure prediction is a beneficial use case, is harmful to the field in how AI affects: pedagogy, research, integrity of knowledge, and quality of care.

As Substitutes for Personnel and Impact on Sociality

AI as staff will affect the patient by allowing a form of stunted parasociality to develop. Cases where this can arise is with people misusing LLM's for self-medicated therapy or AI tools either on the market targeting people in need of mental health support, or AI tools being used by healthcare services (Khawaja, and Bélisle-Pipon, 2023). This would have a negative impact on patients by misguiding them or aggravating their conditions which could pose life threatening risks. Especially, when these models are incapable of acting with any rational or ethical capacity (Sedlakova and Trachsel, 2022). If a patient places excessive confidence into a conversational AI that attempts to act in the capacity of a therapist or similar roles, this could quickly devolve into a form of parasociality. This parasociality would be especially damaging among vulnerable groups that may become dependent upon it for social and psychological needs whilst actively damaging those needs. An example being the following:

> *"The longer our written exchange with ChatGPT went on, the more redundant the information being presented by the chatbot became [70]. It repeated itself over and over again, and any human in the same situation would not likely feel that they were being heard. There were also some contradictions in what ChatGPT had to say that could be very confusing to people living with anxiety or depression."* (Hamdoun, et al 2023, p. 31)

In essence, creating a positive feedback loop that encourages the harmful behaviour. When these LLMs are being provided in the private market as an accessible alternative to clinicians it creates a force currently outside the control of the healthcare sector. It could be argued that the appropriation of scientific signs, and technological progress – in an unqualified manner – is nearing a pseudoscience. Patients' information is also at risk (Khawaja, and Bélisle-Pipon, 2023). For the greatest positive impact in the industry the use of AI should be tightly controlled and not be treated as a substitute for staff.

A positive impact can occur when AI is used as a tool at the behest of the mental health service and not as an agent. There is a tendency among the research to be optimistic but place emphasis on AI being used within the industry as a subordinate. As suggested by Sezgin (2023), the reasoning being that having a human in the loop (HITL) provides oversight to validate the predictions of the model whilst also being able to support the human operator. An example of a positive impact listed is: *"A recent study showed that AI could enhance the accuracy of diagnosis and clinical decisions when combined with expert human evaluation, emphasizing the collaborative nature of AI and doctors."* (Sezgin 2023, p. 2) is citing (Schaffter, et al. 2020). The possibility of users of the technology benefiting is contingent upon the user being aware of the limitations of AI, how to work with it, and not to be reliant on it as a supposed truth machine.

Use in Decision Making and Autonomy

It is most beneficial when there is a HITL of any system where life-affecting decisions are concerned, especially within healthcare. There is a plethora of justifications for having a HITL of processes involving AI; A crucial justification is that it provides oversight for ensuring the correct decisions are being made (Verdicchio, and Perin, 2022). The responsibility is then on said operator as they have the capacity to reason and make ethical and operational decisions that the tool is not trained for or capable of doing. A point here regarding Conversational Artificial Intelligence (CAI) in psychotherapy:

> *"Thus, CAI's strengths can be in the domains and functions that do not require understanding, empathy, and other human features that cannot be reduced to mathematics or statistics. For example, CAI might be better in recognizing patterns and spotting which interventions and techniques are most successful in particular situations. These functions are different from a psychotherapeutic conversation with a human therapist and must be different because CAI is not a rational and moral agent."* (Sedlakova, and Trachsel, 2022, p. 9)

Within healthcare we would want to guarantee that all actions taken are sanctioned and justified. If there were an AI tasked with triaging the consequences could lead to excess mortality a lower quality of healthcare provided. It may implicitly be trained with a bias against those who are disabled or seen as less likely to survive when we would prioritise those populations first to ensure survivability. *"However, chatbots, including ChatGPT, reproduce the biases in the data they are trained on, while also appearing objective and removed from human inconsistencies."* (Hamdoun, et al 2023 p. 29). There are classes of actions which must never be left solely to a machine to decide, for the consequences are too costly and irrevocable. The discourse around the use of machine learning and similar technologies within medicine and healthcare tends to place strong emphasis on human primacy with decision making; it is reasonable to see this as the course of action with the best impact for the sector.

The human operator must know the capabilities and limitations of AI in order to make effective use of it as a tool and to avoid dependency and the degradation of their skills as a healthcare provider. While it is good for there to be a HITL when using Deep Learning, there is still risk if the operator becomes overly trusting of the model's predictions. A dependency upon the technology would atrophy the operator's skills and lead to higher risks of harm being brought to patients (Fiske, Henningsen, and Buyx, 2019). The relationship between user and AI could shift from a qualified operator and model to that of a naive user where there is a reification of the user's belief in the model's "intelligence" and "competence". When used by qualified personnel who use AI with a critical, eye this produces a positive impact for the quality of healthcare.

The "black box" nature of Deep Learning makes it infeasible to determine the "reasoning" behind its decision making, causing observations of subtle forms of bias and influence difficult to identify. There is a great risk of harmful impact upon medicine and healthcare due to opaque variables at play. A common problem with the use of AI has been bias within the training of various models (Khawaja, and Bélisle-Pipon, 2023).

> *"Whatever the nature and shape of the explanation, the goal of XAI is to make the 'black box' of ML systems less 'black' and more transparent, which is undoubtedly a praiseworthy effort. This, however, introduces a further type of problematic cases: to whom to attribute the responsibility for a misdiagnosis in which the attention of the clinician was focused on a particular area of the medical image indicated by the AI (or XAI) system, while in reality the truly critical area connected to the patient's pathology was elsewhere and it was neglected by the clinician precisely because of these enhancements of the AI system that had been introduced to improve the use of AI in medicine? This problem is not new, and it is not limited to medical diagnosis: automation bias is an issue that has been discussed and tackled for decades in any context where humans must take decisions in a highly automatized environment (Skitka et al., 1999)."* (Verdicchio, and Perin, 2022 p. 11)

Whilst this paragraph points out that bias is present, here it feels as though it shifts blame and focus back onto the user of the technology whilst downplaying the explicitly mentioned issue of bias within AI. As this is still, and is likely to remain an issue as long as we use the technology, we cannot give sole trust to this machine to act on its own. According to consequentialist ethics it does not matter the intent or nature of the action so long as the consequences are deemed beneficial. The issue here is that in order to conduct introspection in order to correct undesirable behaviour, it is imperative to discern what calculus is at play in an AI's decision making. As these technologies become further integrated into healthcare systems they will ossify, becoming harder to modify due to the nature of technical debt. This emphasises why the correct behaviour should be verified before being put into production. By reason of these machines' nature, they cannot be treated as equal to an agent.

> *"It is uncontroversial that CAI does not fulfill the standard requirements to qualify as a subject. In the philosophy of mind, mental states or consciousness, autonomy, and intentionality are often postulated in order to acquire the status of an agent or subject (Schlosser 2019).*[5] *CAI is a sophisticatedly developed and effective system for data procession and evaluation that shows or mimic some agent-like features, but it would be highly controversial to attribute the above-listed properties to it."* (Sedlakova, and Trachsel, 2022 p.6)

For best impact, the quality of training must be improved to eliminate bias that is harmful both socially and operationally to the healthcare system and the population it protects.

Conclusion

From this exploration into the question of consequence and possibility within healthcare and medicine at the encroachment of AI, we see that there is still a lot that is unclear about Deep Learning's impact on the industry. Further research is required; it is paramount that it is sober and evermore critical. What can be said is that there is positive impact when there is tight regulation of this technology and AI remains a subordinate tool. There is and can be significant harm should the industry does not tread carefully and fall for the glittering generalities of corporate interest, tech optimism, and unfettered access within markets. The trend within research is that we should use this technology as a tool within a limited capacity thus this must be communicated and mandated for the best impact possible.

References

Fiske A, Henningsen P, and Buyx A. (2019) 'Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy' *Journal of Medical Internet Research*, Available at: https://www.jmir.org/2019/5/e13216 (Accessed: 26/07/2025)

Hamdoun S., Monteleone R., Bookman T. and Michael K., (2023) 'AI-Based and Digital Mental Health Apps: Balancing Need and Risk' *IEEE Technology and Society Magazine*, vol. 42, no. 1, pp. 25-36, March 2023 Available at: https://ieeexplore.ieee.org/document/10063146 (Accessed: 09/08/2025)

Khawaja Zoha, and Bélisle-Pipon Jean-Christophe, (2023) 'Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots' Frontiers in Digital Health. Available at: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1278186 (Accessed: 03/08/2025)

Majnarić LT, Babič F, O'Sullivan S, and Holzinger A. (2021) 'AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity.' *Journal of Clinical Medicine*. Available at: https://www.mdpi.com/2077-0383/10/4/766 (Accessed: 26/07/2025)

Rajpurkar P, Chen E, Banerjee O, *et al.* (2022) 'AI in health and medicine.' *Nature Medicine* Available at: https://www.nature.com/articles/s41591-021-01614-0 (Accessed: 26/07/2025)

Schaffter T, Buist DSM, Lee CI, et al. (2020) 'Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms.' *JAMA Netw Open.* Available at: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2761795 (Accessed: 06/08/2025)

Sedlakova J, and Trachsel M. (2022) 'Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent?', *The American Journal of Bioethics*, Available at: https://www.tandfonline.com/doi/full/10.1080/15265161.2022.2048739 (Accessed: 26/07/2025)

Sezgin E. (2023) 'Artificial intelligence in healthcare: Complementing, not replacing, doctors and healthcare providers.' *DIGITAL HEALTH*. Available at: https://journals.sagepub.com/doi/full/10.1177/20552076231186520 (Accessed: 26/07/2025)

Verdicchio M., and Perin A. (2022) 'When Doctors and AI Interact: on Human Responsibility for Artificial Risks.' *Philos. Technol.* Available at: https://link.springer.com/article/10.1007/s13347-022-00506-6 (Accessed: 26/07/2025)