VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA
Wellington Faculty of Engineering Te Wāhanga Ahunui Pūkaha
Ethics and Sustainability Symposium

The Ethical Considerations of Artificial Intelligence in Clinical Decision Support

Taran John

Abstract

With the explosion in technological innovation facilitating the advent of artificially intelligent systems, specifically clinical decision support, a unique subset of ethical and sustainability concerns arises. Although this technology possesses remarkable potential to revolutionise the healthcare industry, it becomes apparent that an innovative ethical framework must be posited to facilitate integration into the mainstream. Due to the sensitive nature of healthcare, ethical oversights pertaining to incorporation of such technologies would lead to the detriment of its public perception, potentially stigmatising related systems for years to come. By delving into the literature surrounding the idiosyncratic ethical considerations of artificially intelligent clinical decision support in this paper, best practices which seek to mitigate the impact of these concerns emerge. The objective of this work is to assimilate these best practices, which are used in the synthesis of a six principle code of ethics which are as follows: protect healthcare professional authority, ensure technological non-maleficence, cultivate clinical decision support transparency, establish procedures for accountability determination, promote sustainability of artificially intelligence based clinical decision support and encourage equity in the training and deployment of clinical decision support. These principles are then applied to the real world of Watson for Oncology by IBM, to assess the adherence of the product to ethical and sustainability best practices.

Keywords: Artificial intelligence; Healthcare; Ethics; Sustainability.

1. Introduction

Artificial intelligence (AI) is changing the world. Coupled with the skyrocketing edge computing paradigm, AI expansion is facilitating the implementation of technologies that was thought of as only science fiction not two decades ago. From self-driving cars, to virtual assistants, to smart home devices, fields such as transportation, finance and entertainment continue to innovate and implement applications of AI technology at a staggering rate. Embodying this sentiment is an AI model created by Stanford University which was experimentally proven to outperform a board-certified dermatologist pertaining to skin cancer diagnosis, in both accuracy and time efficiency [1]. Requiring only a training data set, instead of years of expensive medical education, the resource requirements between an AI model and a health-care professional (HCP) provide an enticing juxtaposition to a medical field that is struggling with increasingly unsustainable costs and deteriorating outcomes [2]. Although people ranging from computer science experts to politicians are touting AI as a key part of the medical crisis solution, the ethical considerations of such technologies are still hotly debated to this day. It must be noted that this argument does not promote replacement of a HCP with fully autonomous "robot doctors", but instead, recognises the prospective opportunity for transformative cooperation between medical practitioners and AI technology in the form of clinical decision support (CDS). By effectively harnessing a comprehensive evidence-based CDS, significant improvements to clinical diagnosis [3], personalised medicine, [4] and efficiency of operation [5] may become a tangible reality.

1.1. Background

Abundant research pertaining to AI in CDS is present in the literature, endorsed by the claim that it is a "*growing resource of interactive, autonomous, and often self-learning agency*" [6]. For example, traditional machine learning (ML) techniques such as decision trees can be used for breast tumour diagnosis, ensemble learning methods can provide outcome prediction to cancer patients, and Support Vector Machines are used for diabetes diagnosis [7]. More recently, deep learning is growing in popularity in CDS, boasting substantially improved performance in psychiatric, oncology, and optical coherence tomography diagnosis compared to their HCP counterpart [8]. According to recent survey data, the potential annual savings pertaining to the implementation of AI in healthcare can be $150 billion in the United States of America alone, a figure which is catalysing the implementation of this technology into upcoming solutions. For this reason, a recent executive survey ascertained that 69% of life science business have already debuted AI in their solutions and 22% are currently working to implement AI into their technology [7].

1.2. Objective

Despite promising experimental data supporting AI in CDS, proactive ethical considerations must be implemented to facilitate the technology becoming common practice. Therefore, the objective of this work is to review the existing literature pertaining to AI in CDS, identify prevalent ethical complications and best practices, and create an ethical framework based on this research, which is then applied to a real-world technology in the form of a case study. The synthesis of a code of ethics (COE) will not only add to the surrounding literature of this subject, but will also mitigate the risk of inadvertent unethical practices, which in turn could lead to distorted legislation and social rejection, significantly impeding the advancement of such technology into the health-care system.

2. Literature Review

The following section will review the literature surrounding ethical and sustainability issues pertaining to AI in CDS. In turn, ethical considerations that arise through performing critical analysis of each issue will facilitate the synthesis of the COE.

2.1. Ethics and Sustainability Issues of CDS

As detailed in [7], ethics and sustainability issues can be categorised into three different subsections: epistemic, normative and traceability. As such, it would be prudent to present consequential ethical considerations in relation to each of these subsections.

2.1.1 Epistemic Ethical Considerations

Epistemic considerations encapsulate the ethical concerns pertaining to the possibility of inconclusive information or misguidedly trained technologies, or uninterpretable outcomes. Proponents of AI based CDS argue that range of evidence utilised by ML based methods exceed that of a HCP. However, this reasoning fails to account for the fact that an algorithmic determined diagnosis may be inadvertently issued, based on an insufficient amount of evidence. In a similar vein, the conclusions drawn from an AI based CDS are only as reliable and impartial as the data that it has been trained on. For this reason, the susceptibility of an AI based CDS to becoming misguided whilst training presents itself. This sentiment is exemplified in [9], where patients are subject to the possibility of a misdiagnosed heart arrhythmia by an EKG in a smartwatch, based upon either incorrectly captured data due to differences in skin colour, or erroneously calibrated hardware.

Moreover, lack of interpretability is an issue which plagues AI in numerous areas due to the complexity of the mathematical models that are created, and this is no different in the medical field. Therefore, a HCP that is the recipient of a decision made by a CDS technology may not be fully aware of how this outcome was ascertained, nor aware of the training or testing processes in which the model was determined.

2.1.2 Normative Ethical Considerations

Enclosed in the normative ethical considerations are the warranted concerns pertaining to the possibility of transformative effects including invasive patient profiling, or unfair outcomes. Personal autonomy and privacy are fundamental human rights. However, in order to truly harness the full potential of AI based CDS, a vast diverse range of data must be utilised to train the model. This posits a unique contradictory conundrum, as developers and businesses contributing to such technologies have incentive to obtain as much data as possible to create a superior product. Furthermore, these technologies may become susceptible to inadvertent partiality to specific groups of patients due to probabilistic outcomes, whether that be positive or negative. In [10], patients that are perceived to have favourable outcomes due to probabilistic output are prioritised by the algorithm which results in an unintentional discriminatory effect on patients belonging to African and other ethnic minority communities.

2.1.3 Ethical Considerations Pertaining to Traceability Concerns

In a healthcare system involving AI based CDS in cooperation with a HCP, many entities may be involved in collecting and organising the data, developing the model, and utilising the model for a diagnosis. Coupled with the almost uninterpretable mathematical intricacies of an essentially “black-box” model, assignment of ethical responsibility when an unfavourable outcome is reached becomes profoundly complex. The aforementioned viewpoint is corroborated in [11], where the authors reach an unclear conclusion pertaining to the assignment of liability of a negative outcome that was reached by a CDS, thus making future prevention of a similar event difficult.

2.2. Overview of Existing Ethical Frameworks for AI based CDS

Due to the emerging nature of the technology, there are seldom examples of a COE primarily pertaining to AI based CDS in the literature. With that said, however, there exists abundant examples of research relating to the ethics of AI as a whole, a substantial amount of which can be applied to AI based CDS. In [12], the authors compile 84 documents encapsulating the global consensus, through the culmination of government documents such as laws, global welfare organisations such as the World Health Organisation, and COEs from technology businesses such as Microsoft. From these, a worldwide convergence around the five ethical principles of transparency, justice and fairness, non-maleficence, responsibility, and privacy were ascertained.

3. Code of Ethics

The synthesis of the following COE encompasses the five major ethical principles surrounding AI technology, and adapts them to the application of CDS, which strives to address the aforementioned epistemic, normative and traceability concerns.

- **Principle 1: Protect HCP Authority.**
  An HCP must always have complete oversight of the CDS technology, and has full authority pertaining to all decisions made during the clinical process.

- **Principle 2: Ensure Technological Non-maleficence.**
  All entities involved in the manufacturing and deployment of AI based CDS must ensure that patient safety is the foremost concern, and to ensure that they are not subjected to decisions made by technology that has not been proven to be efficacious in both an experimental and clinical environment.

- **Principle 3: Cultivate CDS Transparency.**
  Entities responsible for the development of an AI based CDS, must ensure that the process in which a decision was made by the system is interpretable to a trained HCP. Moreover, transparency pertaining to passive data being collected from a patient must be present, and be collected only with their express written consent.

- **Principle 4: Establish Procedures for Accountability Determination**
  Mechanisms must be implemented into the system to ensure that the correct entity, from manufacturer to clinician, is held accountable for unfavourable outcomes due to AI based CDS.

- **Principle 5: Promote Sustainability of AI based CDS**
  Manufacturers of CDS apparatus must implement mechanisms which facilitate frequent updating of the system, and must immediately rectify ineffective technology.

- **Principle 6: Encourage Equity in Training and Deployment of CDS**
  CDS technologies must strive to be effective for every patient, regardless of ethnicity, gender, age, or size. Furthermore, these tools should be available for use in multiple languages in institutions across the globe.

3.1. Principle 1: Protect HCP Authority

Addressing perhaps the most prevalent criticism of AI in CDS being loss of human autonomy, this principle seeks to ensure that a HCP has full oversight and authority of the clinical process. This will substantially mitigate the risk of significantly incorrect decisions made by the system as a result of epistemic and normative issues. Drawing inspiration from the responsibility and traceability ethical concepts, this principle promotes the idea that the experienced, human element of clinical decision making by an HCP must remain a fundamental component to this process, and removing the ambiguity surrounding accountability of unfavourable outcomes.

3.2. Principle 2: Ensure Technological Non-maleficence

Just as a HCP is obligated towards maintaining the ethical principle of non-maleficence, so must any individual, group or business related to the manufacturing, experimentation, and deployment of AI based CDS. This includes eradicating any potential for unproven, experimental CDS technologies to be consulted under any circumstances. Upholding this principle should reduce the foreseeable epistemic risks of misguidedly trained models, and the normative risk of unfair outcomes.

3.3. Principle 3: Cultivate CDS Transparency

Seeking to integrate the ethical principle of transparency, this principle states developers and manufacturers can only deploy AI based CDS technologies into a clinical setting, when they can be fully understood by a trained HCP, the patients involved, and a regulatory entity, to address epistemic issues. Furthermore, in order to combat the normative issue of invasive patient profiling and uphold the ethical principle of privacy, the overseeing HCP must obtain the express written consent of the patient to utilise their data for training or diagnosing purposes.

3.4. Principle 4: Establish Procedures for Accountability Determination

Through integration of processes which appropriately assign liability to the entity accountable for unfavourable outcomes, the ethical principle of responsibility is upheld, and the ethical concern of traceability is addressed. Assigning blame to the appropriate party will also facilitate improvements in the concerned area, in order to reduce the risk of this outcome from happening again.

3.5. Principle 5: Promote Sustainability of AI based CDS

AI based CDS manufacturers and developers should continuously be monitoring and seeking to improve deployed technologies, such that if any epistemic or normative concerns should arise, it will be adjusted in a timely manner. Furthermore, in-line with the ethical principle of responsibility, institutions seeking to integrate AI based CDS into their practice must only do so if they are in possession of the resources needed to immediately rectify ethical breaches that might arise.

3.6. Principle 6: Encourage Equity in Training and Deployment of CDS

Evidence of existing AI based CDS technology breaching the ethical principle of justice and fairness is present throughout the literature, with warranted epistemic concerns of misguided data and normative concerns of unfair outcomes. This principle aims to combat this by promoting training and testing, utilising diverse data sets, and fostering the idea that manufacturers and developers must keep in mind all types of patients.

4. Case Study

The advent of numerous emerging AI based CDS technologies provides many opportunities for the demonstration of the aforementioned COE. However, it would be prudent to apply it to a pioneering technology that is used worldwide, such as Watson for Oncology (WFO). With the assistance of the premier oncologists from the Memorial Sloan Kettering Cancer Centre (MSKCC), WFO is an AI based CDS created by IBM Corporation [13]. Trained for over four years employing 100 years' worth of United States clinical data and experience, WFO provides patients with diagnosis support and tailored chemotherapy treatments. WFO candidate treatments are partitioned into three subdivisions: "green" which is recommended, "yellow" which could be a potential alternative, and "red" which is not recommended due to obvious evidence against its use in this case. Furthermore, a multi-disciplinary team (MDT) team comprised of oncologists, surgeons and other medical disciplines, are also tasked with the objective of maintaining consistency between WFO-clinician cooperation, over different regions and demographics. The MDT discusses the benefits and detriments of each possible treatment, and based upon these results, categorise these into "concordant" or "discordant" consensus opinions, respectively.

4.1. Principle 1: Protect HCP Authority

The manner in which WFO was implemented into clinics adheres to this principle, as it prioritises HCP autonomy. Clinicians are offered evidence-based recommendations based upon the data provided by a patient, ultimately allowing them to make an informed decision whether to act in accordance or disregard the suggestion made. Additionally, the presence of an MDT of medical experts which routinely critique the system ensures that the human element of healthcare remains foremost, mitigating the risk of epistemic or normative concerns leading to clear misdiagnoses.

4.2. Principle 2: Ensure Technological Non-maleficence

Abundant criticisms surrounding the efficacy in correctly diagnosing patients in a clinical environment, eliminating the possibility of claiming that WFO adheres to this principle. In particular, the claim that "*IBM has discovered that its powerful technology is no match for the messy reality of today's healthcare system*" is prevalent [14]. Moreover, IBM has been accused of releasing a product that has only been proved to be effective in controlled, experimental environments, and which is not ready to be translated to a clinical setting.

4.3. Principle 3: Cultivate CDS Transparency

Interpretability of this technology is shown to be severely lacking, therefore causing misalignment between the system and this principle. Accusations of the system consistently providing purposeless and even dangerous recommendations are numerous, with presiding HCP uncertain of how these decisions were made [14]. It must be noted that privacy was not breached with the training of this model. However, more healthcare institutions will need to provide consensual patient data to facilitate a more efficacious model.

4.4. Principle 4: Establish Procedures for Accountability Determination

Given that the overseeing HCP retains full autonomy of overall decision making, coupled with the willingness of IBM to address prevalent criticism of WFO [15], it becomes clear that this system adheres to this principle. To elaborate, a HCP is met with an evidence based decision as an output of the CDS, in conjunction with a "concordance score" which portrays the opinions of other clinicians, and can therefore make an informed diagnosis. Therefore, accountability rests solely upon the HCP, who can choose to accept, reject or partially factor the output of the CDS in the decision.

4.5. Principle 5: Promote Sustainability of AI based CDS

Currently, the sustainability of WFO adheres to this principle, but is heavily predicated upon the clinical success of the product. With over $4 billion invested in WFO, IBM has the resources necessary to successfully integrate the system into clinics and hospitals world-wide [14]. WFO additionally contains mechanisms which facilitate regular updates to the model, and IBM is continually attempting to improve the model. With that said, without tangible clinical evidence of efficacy, it is likely that investors will look elsewhere, and the WFO project will be shelved.

4.6. Principle 6: Encourage Equity in Training and Deployment of CDS

Perhaps the most prevalent criticism surrounding WFO is the lack of diversity in the training data, and for this reason, any claim that it is an equitable system is unfounded. In fact, after the deployment of WFO in China, issues with concordance and presence of unfavourable outcomes were more prevalent with Chinese patients compared to their Western counterparts [15]. Moreover, since the system had primarily been trained on data provided by the MSKCC, claims that the system was trained only on data obtained from those of a wealthy demographic also surfaced.

5. Conclusions and Recommendations

AI has been proven to have unique potential to transform the healthcare industry through CDS. However, despite possessing a unique set of ethical, normative and traceability requirements, there remains a lack of ethical frameworks surrounding AI based clinical resources. Through identification of best practises in the literature, this work has synthesised a tangible COE, and has provided an example of its functionality by applying it to a prevailing real-world AI based CDS, WFO. Inescapably, there are some limitations to this approach. Firstly, this work was restricted to literature that was published in English, meaning that potentially critical documentation written in another language such as Spanish or Mandarin was not taken into account when surveying the literature. Additionally, academic literature tends to lag behind state-of-the-art technology, giving the possibility that potentially major ethical concerns have arisen but is not yet present in the research. Future work should involve the survey of documentation written in other prevalent languages and multiple real-world technologies, discussing patterns pertaining to their proficiency in implementing mechanisms of reducing ethical concerns. Perhaps the predominant recommendation to the engineering industry regarding AI based CDS is to ensure new technologies in the field are proven to be efficacious in a clinical environment before overpromising on a product that cannot deliver without significant ethical shortcuts being taken. Instead, companies should ensure adherence to a robust COE, rather than potentially providing a truncated product which dismantles public perception of a technology that could transfigure the healthcare industry.

References

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[2] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.

[3] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O’Donoghue, D. Visentin, et al., “Clinically applicable deep learning for diagnosis and referral in retinal disease,” Nature Medicine, vol. 24, no. 9, pp. 1342–1350, 2018.

[4] C. Barton, U. Chettipally, Y. Zhou, Z. Jiang, A. Lynn-Palevsky, S. Le, J. Calvert, and R. Das, “Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs,” Computers in Biology and Medicine, vol. 109, pp. 79–84, 2019. 8

[5] A. Nelson, D. Herron, G. Rees, and P. Nachev, “Predicting scheduled hos- pital attendance with artificial intelligence,” NPJ digital medicine, vol. 2, no. 1, pp. 1–7, 2019.

[6] L. Floridi, “Ai opportunities for healthcare must not be wasted,” Health Management, vol. 19, no. 2, 2019.

[7] J. Morley, C. C. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo, and L. Floridi, “The ethics of ai in health care: a mapping review,” Social Science & Medicine, vol. 260, p. 113172, 2020.

[8] A. V¨a¨an¨anen, K. Haataja, K. Vehvil¨ainen-Julkunen, and P. Toivanen, “Ai in healthcare: A narrative review,” F1000Research, vol. 10, no. 6, p. 6, 2021.

[9] R. Hailu, “Fitbits and other wearables may not accurately track heart rates in people of color,” Stat News, 2019.

[10] C. Garattini, J. Raffle, D. N. Aisyah, F. Sartain, and Z. Kozlakidis, “Big data analytics, infectious diseases and associated ethical impacts,” Philos- ophy & technology, vol. 32, no. 1, pp. 69–85, 2019.

[11] E. Racine, W. Boehlen, and M. Sample, “Healthcare uses of artificial in- telligence: Challenges and opportunities for growth,” in Healthcare man- agement forum, vol. 32, pp. 272–275, SAGE Publications Sage CA: Los Angeles, CA, 2019.

[12] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of ai ethics guidelines,” Nature Machine Intelligence, vol. 1, no. 9, pp. 389–399, 2019.

[13] Z. Jie, Z. Zhiying, and L. Li, “A meta-analysis of watson for oncology in clinical application,” Scientific reports, vol. 11, no. 1, pp. 1–13, 2021.

[14] E. Strickland, “Ibm watson, heal thyself: How ibm overpromised and underdelivered on ai health care,” IEEE Spectrum, vol. 56, no. 4, pp. 24–31, 2019.

[15] W. Liu, X. Shi, Z. Lu, L. Wang, K. Zhang, and X. Zhang, “Review and approval of medical devices in china: changes and reform,” Journal of Biomedical Materials Research Part B: Applied Biomaterials, vol. 106, no. 6, pp. 2093–2100, 2018.