The Ethical Analysis of Introducing Artificial Intelligence into Healthcare

A Research Paper

in STS4600

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Biomedical Engineering

By

Michael Hughes

May 8th, 2020

Approved by Prof. Kent Wayland, Department of Engineering & Society

Introduction

The internet is a millennial, the iPhone is a pre-teen, and Google would be in college right now. The technological advancements of the past ten, twenty, thirty years have been astounding. Not long ago the internet was the most profound advancement of the time, and now we're knocking on the door of autonomous technology. Self-driving cars are all the rage now and they, along with other autonomous technologies, are what big technology companies like Apple, Google, and Uber are focusing on. It is generally considered that advances in technology are inherently good and make society better, however, as we continue to push boundaries and advance technology towards autonomy, will this always be true? Stephen Hawking, one of history's most influential scientists, once said "[Artificial Intelligence] is likely to be either the best or worst thing to happen to humanity." (Hawking, 2015) What is the cost of accomplishing more as a society through advancing technology? Would advances into an area such as healthcare actually diminish society's overall quality of life? Current physicians provide a human touch to patient care making patients feel more comfortable in the exam room. This is something AI would not necessarily be able to provide. People in need of care check into the healthcare system at a vulnerable point looking not only for help, but also for compassion. If continued without the proper knowledge, automation could eliminate several important aspects of healthcare that only come with human interaction.

There are two sides to the coin that is the introduction of artificial intelligence into an area such as healthcare. On one hand, these programs could possibly improve accuracy and efficiency of patient care, but on the other hand, they could negatively change the way patient care is performed. Would implementing such programs do more good or more harm? In order to shed light on the matter, this thesis will explore the ethical issues involved in AI in healthcare

and help better understand how the quality of life for society, specifically healthcare patients, will be affected by the potential introduction of AI into the healthcare space. To do so, two ethical schools of thought will be used to add to the knowledge base of ethics and AI, so that when the opportunity for AI to be introduced into healthcare does come, society will know that patients, even in their most vulnerable times, are receiving the highest quality care possible.

The History and Future of Automation

To be automatic is to have the capability of starting, operating, or moving independently. Automation is defined as a technique, method, or system of operating a process by automatic means, reducing human intervention.

Despite the fact that artificial intelligence has only recently made its charge to the forefront of technological development, the notion of automation has been around for decades. A massive advancement in automation came during the Second World War when a British computer scientist named Alan Turing built a machine to crack the German enigma code and detect the movements of the opposition (Ray, 2019). Turing was also the one to coin what is called the imitation game. This 'game', as Turing describes, is one that helps to define an intelligent machine by determining whether said machine could converse with a human without the human's knowledge that he or she was speaking to a machine. This view of intelligence is how most people today view artificial intelligence. Since Turing's great innovation in the 1940's, several AI Winters, periods of reduced funding and interest in AI research, have hindered advances in artificial intelligence. In recent years, the growth of technology super companies such as Apple, Google, Uber, Tesla, and Microsoft have moved artificial intelligence forward leaps and bounds. More and more, AI is becoming what Turing predicted. Behind the anonymity of a screen, many internet users would be unable to identify if the 'person' in the chat with them

was really a person or a customer service bot. Companies such as Google and Tesla are implementing such advanced AI that they can drive cars themselves and make decisions on where to turn, how fast to go, and when to swerve or break.

The technology of AI is rapidly advancing and could quickly be a larger part of the healthcare space than it already is. Here, autonomous healthcare agents are described as programs or AIs that perform a function in the healthcare space that would have otherwise been performed by a person. This also means that they have a wide variety of current and potential applications. Presently, algorithms and autonomous programs have limited use in healthcare, and the ones that are in use are mainly being used for patient data collection or very limited diagnostic tools (Ash et al., 2019). In the future, AI has the potential to be used in virtual human conversation tools for collecting patient information, patient diagnostic tools, patient data tracking with wearable tech, and administrative activities (Davenport & Kalakota, 2019). The roles of AI in healthcare have the potential to become a more prominent fixture in the sector. When AI is beginning to be introduced into healthcare, society needs to be ready to answer the ethical questions this technology introduces. This thesis will analyze AI in healthcare as well as other areas of technology in order to address the most relevant ethical concerns.

Two Schools of Ethical Thought

In regards to ethical theory there are two that will be used in this paper: utilitarianism and deontological (duty) ethics.

Utilitarianism was created by Jeremy Bentham in the late 18th-century and was centered around the idea of utility. A concept commonly used in economics, utility is defined as a state of being useful, profitable, or beneficial. The way Bentham saw it, the utility of an action was

centered around his value theory of hedonism that described pleasure as the only thing that is good in itself. This theory is what Bentham based his ethics around deeming actions as good or bad based on the measure or pleasure they produced. Actions which provide pleasure and avoid pain are good, and that which provide pain or reduces pleasure is bad (van de Poel & Royakkers, 2011). However, what may be pleasurable to one certainly might not be pleasurable to all. That is where his notion of a moral balance sheet is introduced. Bentham believed that in deeming if an action was good or bad all persons must be taken into account. The last key piece to Bentham's ethical theory was actually introduced by his student John Mill in the mid-19th-century. Mill introduced that notion that actions could not just be weighed as binary good or bad, but also must be associated with a weight or quality measure (van de Poel & Royakkers, 2011). So, weighing the good and the bad, the actions that produced the largest amount of good (pleasure) were deemed the best actions. In order words, actions that produces a lot of pleasure were preferred to those that produced only a small amount of pleasure. This form of utilitarianism is called act utilitarianism in which consequences of actions are specifically judged, and will be the primary focus of this paper along with the next ethical theory.

Deontological ethics, created by Immanuel Kant in the late 18th-century, holds that an action is morally right if it is in agreement with a moral rule. This rule is to be applicable in itself, independent of the consequences of the action that is under question (Powers, 2006; van de Poel & Royakkers, 2011). The way in which Kant differed from Bentham is that he thought these rules could not be based on happiness or pleasure. He considered happiness to be subjective, something that changes from person to person or even within one person over time. Kant established his theory of ethics around the idea that creating a moral norm or set of rules would allow for a reference point for actions to be judged. Then, because these would be considered the

norm, everybody would agree on them and it would be everybody's duty to adhere to them (van de Poel & Royakkers, 2011). The way in which Kant decided these rules would be created was such that all rational people could be capable of judging whether an action is morally right. In simpler terms, these rules should be so clear as to not cause any confusion as to what is right or wrong.

Current Knowledge Regarding Ethical Autonomous Healthcare Agents

Currently, there is no clear consensus in the field as to the ideal theory to understand the ethical concerns pertaining to the use of intelligent healthcare agents (Anderson & Anderson, 2008). Literature in the field shows that the two ethical theories used in this paper contrast each other in several ways when applied to the use of AI in healthcare. Utilitarianism would posit that whichever action provides the most total good onto society would be the correct decision (Driver, 2014). Deontological ethics, on the other hand, would only mark a decision as correct if it followed some established moral rule, regardless of the overall consequence (Alexander & Moore, 2007). Utilitarianism is society-centered, focusing on the well-being of society as a whole, while deontology is patient-centered, focusing on the well-being of each individual, an issue that integrating AI into healthcare may introduce (Mandal et al., 2016).

Methods for Data Collection

This research paper aims to build on the current work and literature done regarding the ethical considerations surrounding AI and specifically AI in healthcare. In order to do so, an analysis of the ethical theories described above will be combined with the extensive investigation of the ways in which technologies like AI effect the way the ethical theory is built to govern it, and vice versa. The data collected to address the research question will mostly be through the

documentary research method in which each source will be certified for authenticity, credibility, and representativeness. After collecting these data, the data will be processed by doing thorough background research on the source and author to better understand the perspective the document is written from. Once the data are processed, any trends and statistics that can be extracted from the document will be taken and extrapolated to show how such ethical theories can be applied to AI in healthcare. Doing so, will shed more light on AI and its role in healthcare and provide insight to the research question.

How the Ethical Questions Surrounding AI in Healthcare Could be Evaluated

Using utilitarianism, the paper will attempt to determine if the consequences brought on by the introduction, or further advancement, of AI in the healthcare space produce a positive or negative reaction to those effected. The thesis will attempt to account for all effected parties as well as the associated weights, positive or negative, with each party to create an accurate ethical balance sheet and assess the ethical standing of the AI and the questions surrounding it. At the same time, the paper will use deontological ethics to assess if AI upholds the same moral rules established for those in the healthcare field. Additionally, it will be important to note if any other moral rules will need to be created for the use of AI in healthcare, and then if the AI will be able to uphold these rules at all times.

One-way experts see AI moving forward in the healthcare field is through Ambient Intelligence or Ambient Clinical Intelligence, AmI or ACI. Ambient Intelligence is the latest vision of the internet of things (IOTs) movement. Using AmI people are empowered through a digital environment that is aware of their presence and context and is responsive to the user (Riva, 2003). In order to envision what AmI in healthcare or ACI is like, picture a typical exam room with a tv display in it along with a set of microphones. These microphones would use

familiar speech recognition technology to provide the information the clinician desires as well as track the patient's visit. By integrating AI, machine learning, and cloud computing, the ACI system is able to provide diagnostic guidance as well as clinical intelligence. That includes highlighting potentially overlooked diagnoses based on patient history and symptoms, as well as possible drug interactions and recommend alternative medications (Ash et al., 2019). As noted by the Information Society Technologies Advisory Group, such AmI environments will be all around us, will be sensitive to presence of living creatures and support their activities. The environment will 'remember and anticipate' behavior in the new virtual world (Riva, 2003). With the integration of interactive technology into our environment, privacy concerns arise.

Currently, individuals are concerned with low-level AIs such the Google Home or Amazon's Alexa keeping track of their whereabouts or listening in on their conversations. AI applications can be and have been used to identify, and thereby track, individuals across different devices, in their homes, etc. (*Artificial Intelligence | PI*, n.d.). If things such as these are a concern to the public, how might society react to an AI present the doctor's office? Interestingly enough, the doctor's office is usually where society draws the line in regards to where they are concerned with their whereabouts being tracked. This is a concept Helen Nissenbaum coined as contextual integrity. In short, Nissenbaum describes the notion that privacy is neither the right to secrecy nor the right to control, but a right to appropriate flow of personal information. Central to contextual integrity are the norms established for the flow of information (Nissenbaum, 2009). When it comes to medicine, people are much more willing to have their information be known by the required parties such as physicians and nurses. It is considered a norm to keep this flow of information strictly between the patients and the healthcare providers. This is why many people often agree with the informed consent agreements presented to them by their physician or

medical researcher. So, while ACI may seem to present an ethical concern regarding a patient's privacy, when looked into further it would appear not. From a deontological perspective, this ethical question shares a similar view as utilitarianism. It is generally accepted that invasion of one's privacy is unethical. There are social rules and laws in place such as the Gramm-Leach-Bliley Act, the Health Insurance Portability and Accountability Act (HIPAA), and the Children's Online Privacy Protection Act that protect society from an invasion of their privacy. However, as just mentioned, when patient's check in to the doctor's office, many actually prefer their physician to know their information. ACI and AmI environments would still need to protect from an invasion of privacy from outside the doctor's office under HIPAA, however, inside the exam room it would appear ethical for this technology to track health information under these two ethical theories.

Researchers are facing many hurdles when attempting to develop an autonomous healthcare agent. First, they are finding it difficult to get the program to determine what is deemed as good and bad and the quality of each when framing the program using utilitarianism. Ideally, the program would be able to do this without any human input, however, introducing human input to establish this set point also introduces elements of framing bias, stereotypical bias, and availability heuristic amongst other specific biases typical to human decision making. One reason for delegating such ethical questions to an AI, along with improving accuracy and efficiency, is to remove these certain biases. If the program cannot establish a reference point for good and bad without the programmer doing so, then that reason for an AI being used is nearly defeated (Anderson & Anderson, 2008).

Researchers have also attempted to use duty ethics to develop an ethical artificial intelligence to be used in healthcare. Many programmers believe that a single-principle, absolute

duty ethical theory would not be appropriate as they lack the ability to capture the necessary complexity of the ethical issues present (Anderson & Anderson, 2008). Instead, many researchers are turning to theories that consist of several primary duties, such as that of Ross' *prima facie*, in which a set or list of duties is created as opposed to a single rule (Ross & Stratton-Lake, 2002). In establishing a program to follow such a theory, a set of rules need to be established for the AI to follow, as suggested by the Kantian (Deontological) theory. The same way rules are established to ensure humans follow an ethical path, so is done for the program (Powers, 2006). However, as mentioned previously, requiring human input to provide a base line for the program introduces the several biases listed before.

The largest area in which developers are attempting to make an impact with the introduction of AI is in that of diagnosis. Using AI and cloud-based computing, "AI physicians" would have access to limitless amounts of data kept on a patient as well as information around the globe to compare a specific patient case too. Currently, the introduction of AI into patient diagnosis is seen to be primarily as a physician's assistant similar to the ACI system mentioned previously. Early AI systems are currently being used to help in the diagnosis of cancer, neurology based diseases or disorders such as Parkinson's and strokes, and cardiac abnormalities (Jiang et al., 2017). Right now, a clinician is still the center of the decision-making process; however, as AI becomes a more prominent part in healthcare as expert predict, will they be the ones making decisions? The ethical concern that comes with scenarios such as this are in regards to accuracy and then accountability.

In terms of accuracy, it would come down to the numbers. If the program correctly diagnosed a significantly larger number of people than it misdiagnosed and thus saved or benefitted more lives than it hindered, from a utilitarian perspective AI could be ethical here. The

weighting of costs and benefits would come into play; however, that analysis would be better suited when the time comes. Deontologically, AI diagnostic errors would be viewed as unethical. If it is known that the program has an error rate, it would be unethical to subject patients to the chance that the AI could be wrong. It is generally accepted that it is morally wrong to leave the well-being or life of even one person up to chance. A similar question is being asked now as self-driving cars are being developed. Self-driving cars will inevitably crash and thus have the potential to injure people. In such a scenario, an expansion of the famous trolly problem, the car has to make a decision: Avoid a serious collision and save the passenger in the car, potentially killing people outside the car or stay the course saving the lives of pedestrians while potentially killing the passenger (Nyholm & Smids, 2016). Interestingly enough, most people are utilitarian, meaning they would choose to save the most people possible, until they are the passenger in the car whose life is on the line (Ackerman, n.d.).

There are also ethical concerns regarding accountability. As aforementioned, it is inevitable for the AI to make a mistake. Unfortunately, in both healthcare and autonomous vehicles, when mistakes are made lives are usually on the line. When serious crashes or fatal misdiagnoses occur, people are disposed to want to find some person or persons to be held responsible (Nyholm & Smids, 2016). This is because people expect the person held responsible make the situation right. Society has actually developed a way to ensure that the correct people are held accountable and that the wrong is made right. This is called the legal system. It is difficult, however, to use the legal system to hold an AI accountable isn't it? You can't drag an AI in front of a judge to face justice. So, the question arises, who is to be held responsible for serious missteps an AI makes? This question is a difficult one as it causes problems in two main areas of the legal system: causality and compensation. To analyze accountability of AI let's start

with a more familiar example: Person A shoots and kills Person B with a gun. Using familiar legal processes, Person A would be held accountable for the death of Person B because Person A's decision to pull the trigger is the *greatest cause* of Person B's death (Bartlett, 2019). When looking at AI, it is difficult to say whether the system's decision making is the 'greatest cause' in the death of a passenger in a self-driving car or a patient in a misdiagnosis as opposed to the developer, so we continue analysis. Due compensation is also important when it comes to holding someone accountable. Victims of an accident or crime wish to be compensated for what they lost, whether that be property, health, or even a loved one. This is obviously a problem when the 'someone' being held accountable is an AI with no monetary holdings or anything to offer thus pushing accountability away from the AI itself and towards someone who can provide due compensation. This push towards a human, or group of humans, begins to point at the developers of the AI. The AI developers, the person or persons who directly shape the programming of the AI, are the group that is the *greatest cause* for the actions of the AI even if the AI itself carried out the actions. Another familiar analogy is that of a misbehaving child who destroys property in a store. It would be reasonable to expect the parents of the child to be held accountable for the damages despite not directly causing them (Bartlett, 2019). In today's world AI is juvenile. Perhaps, one day AI will be able to face a judge and a jury, will be able to be punished for its actions and provide compensation. However, until that day comes, an AI's actions must be tied to those who created it. The analysis done on accountability did not use the two aforementioned ethical theories, utilitarianism and duty ethics, however, shows that evaluating difficult ethical questions should be done from multiple perspectives.

Conclusion

In light of the rapid development of AI, researchers and ethicists have flocked to the drawing board to hopefully understand how this advanced technology will affect society's way of life. This thesis has analyzed the ways in which an introduction of AI into healthcare could change the landscape of the healthcare system and whether those changes are ethical for society as a whole and for individuals. Utilitarianism allows for the changes that AI causes to be weighed on a scale of good and bad. The scale seems to indicate that introducing AI would be ethical due to the sheer number of patients that could be benefitted; however, the potential to lose just one human life could tip the scales the other way. Further research into the weight of a human life would help society better understand if AI in healthcare would be a net benefit or harm. Deontological ethics analyzes the changes AI would make through the perspective of societal norms. It is considered an ethical rule to not do harm unto others willingly, so the knowledge that introducing AI into healthcare could potentially cause harm or injury to others would make it unethical. The arguments presented here, as well as the cases they help to address, were chosen as they appeared to be the most prevalent in regards to AI in healthcare, as well as AI in general. Questions regarding privacy, accuracy and accountability will continue to be the largest concerns during the development of AI in healthcare and further analysis will shed light on these issues. Even though work is being done to better understand the ripple effect regarding changes in healthcare, we still do not understand every perspective from which the ethical concerns addressed here can be viewed. It is necessary to continue to view these ethical questions through a variety of different ethical lenses, such as virtue ethics, care ethics and ethics of rights for example, in order to better understand the impact of AI in the healthcare space.

Reference List:

Ackerman, E. (n.d.). People Want Driverless Cars with Utilitarian Ethics, Unless They're a Passenger—IEEE Spectrum. *IEEE Spectrum: Technology, Engineering, and Science News*. Retrieved March 5, 2020, from https://spectrum.ieee.org/cars-that-think/transportation/self-driving/people-want-driverless-cars-with-utilitarian-ethics-unless-theyre-a-passenger

Alexander, L., & Moore, M. (2007). Deontological Ethics. In *Stanford Encyclopedia of Philosophy* (Winter 2015). https://stanford.library.sydney.edu.au/archives/win2015/entries/ethics-deontological/#PatCenDeoThe

Anderson, M., & Anderson, S. L. (2008). Ethical Healthcare Agents. In M. Sordo, S. Vaidya, & L. C. Jain (Eds.), *Advanced Computational Intelligence Paradigms in Healthcare—3* (Vol. 107, pp. 233–257). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77662-8_10

Artificial Intelligence | PI. (n.d.). Retrieved March 4, 2020, from https://privacyinternational.org/learning-topics/artificial-intelligence

Ash, M., Petro, J., & Rab, S. (2019). *How AI in the Exam Room Could Reduce Physician Burnout*. 5.

Bartlett, M. (2019, April 5). Solving the AI Accountability Gap. *Towards Data Science*. https://towardsdatascience.com/solving-the-ai-accountability-gap-dd35698249fe

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Driver, J. (2014). The History of Utilitarianism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/

Hawking, S. (2015, March 30). Zeitgeist 2015 Conference, London, England.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Mandal, J., Ponnambath, D. K., & Parija, S. C. (2016). Utilitarian and deontological ethics in medicine. *Tropical Parasitology*, *6*(1), 5–7. https://doi.org/10.4103/2229-5070.175024

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.

Nyholm, S., & Smids, J. (2016). The Ethics of Accident-Algorithms for Self-Driving Cars: An Applied Trolley Problem? *Ethical Theory and Moral Practice*, *19*(5), 1275–1289. https://doi.org/10.1007/s10677-016-9745-2

Powers, T. (2006). Prospects for a Kantian Machine. *Intelligent Systems, IEEE*, *21*, 46–51. https://doi.org/10.1109/MIS.2006.77

Ray, S. (2019, December 15). *History of AI*. Medium. https://towardsdatascience.com/history-of-ai-484a86fc16ef

Riva, G. (2003). Ambient Intelligence in Health Care. *CyberPsychology & Behavior*, *6*(3), 295–300. https://doi.org/10.1089/109493103322011597

Ross, W. D., & Stratton-Lake, P. (2002). *The right and the good* (New ed.). Clarendon Press.

van de Poel, I., & Royakkers, L. (2011). Ethics, Technology, and Engineering. *Wiley-Blackwell*, *1st ed.*, 378.