SOCIO-TECHNICAL DESIGN FOR CLINICAL DECISION SUPPORT ALGORITHMS

Sven de Witte
TU Delft
June 2021

Abstract

Clinical decision support systems is a collection name for a lot of Artificial Intelligence systems used in healthcare. These systems are designed to help health workers make decisions faster and make the healthcare environment as a whole more efficient. Decisions made by these systems often weigh heavy on the ethical side giving advice on what kind of care a patient receives. The ethics of these decision mean it is even more important for these systems to act fair and unbiased towards all patients. This however is not always the case though. This paper will explain and dissect the issues of unfairness within clinical decision support systems and will compare and adapt different socio-technical design methodologies to form a advice on designing for fair clinical decision support algorithms.

1 Introduction

In recent years Artificial Inteligence (AI) has been introduced in a large number of domains. Healthcare is one of them, where AI is being used for a wide spread of uses such as diagnosing of diseases, classifying genes and predicting outcomes of certain procedures[16]. By introducing AI into the environment of healthcare a few of the common socio-ethical issues AI struggles with are also introduced.

There are multiple papers on the topic of ethical AI systems and AI in healthcare, where a paper like that of Mittelstadt categorises the different issues[15]and a paper like that of Morley [16] looking at the ethics of artificial intelligence in healthcare. Or papers like that of Leslie[14] focusing of inequality of AI in healthcare especially in the Covid times. On the other side there are papers such as of Selbst[19], Aizenberg[3] and Baxter[4] that go into how technical systems with high social and ethical functions should be improved and designed on the broader scale. Presenting methodologies and guidelines for designing these socio-ethical AI systems.
Many of the papers show the presence of these issues and the importance of the awareness for these issues. What is missing are more focused papers on the different categories of algorithms in different environment, trying to prevent and solve the specific problems present in those environments.

The healthcare system has many different uses for AI that could be and should be explored. This paper will focus on a smaller subset of AI uses within healthcare, mainly the clinical decision-support systems. These are systems which help healthcare workers make decisions by providing suggestions or by predicting risks or outcomes. While many algorithms could fall under this category this paper will focus on algorithms that are aimed at limited resource admission such as available space, time and medication. Not only do these algorithms have a lot of ethical decisions to make, but also a lot of the newer technologies and AI within healthcare fall within this category. An example of such algorithm is for instance an algorithm deciding which patient to send to or keep on the Intensive Care Unit(ICU), which has limited space available, based on predictive algorithms predicting risks and chances of survival trying to keep the most people alive.

The issue of unfair outcomes covers cases such as algorithms that learn to prioritise patients it predicts to have better outcomes which turns out to have discriminatory effects on people of color[10]. This is a big social-ethical issue and solving or reducing this would not only benefit the healthcare environment but also the field of AI. The problem with trying to solve this issue is that it is not just a purely technical problem or a purely social problem and thus requires something more than a technical or social solution. This is where socio-technical design and thinking is necessary. Socio-technical design approaches aim to design in a way that considers the human, social, organisational and technical factors of a socio-technical system.

This paper serves as a guide into the issue of unfair outcomes of clinical decision-support AI and the socio-technical design approaches that can help address this issue. At first the most common examples of unfairness within clinical decision-support systems will be listed and explained. Explaining the impact that these issues have on healthcare and the different ways in which unfairness represents itself. Following this a deeper insight will be given into the cause of these examples of unfairness. Diving deeper into the underlying problem from a social and a technical point of view. The third section of this paper will focus on the already existing socio-technical design approaches already used in other fields or that are explained in other papers. At last this will all be linked together to give a guideline into the socio-technical design approaches that can be used to design and improve clinical decision-support systems.

2 Methodology

To answer and research the different topics mentioned in the introduction a literature study will be done with papers of computer science, healthcare and social fields. Sources for this literature will be papers and articles drawn from Google Scholar, PubMed(National Library of Medicine), Nature which is a platform for peer reviewed research papers and HealthAffairs which also provides peer reviewed research papers but aimed at health policy issues.
Keywords in the search terms will be: Bias, Fairness, Discrimination, unfair, Socio-technical, sociotechnical, socio-ethical, Clinical decision support algorithm, Clinical decision support system, Healthcare, Health, Algorithm. Besides papers and articles found in this way the provided papers by the supervisor will be used as well as any relevant papers from references within the papers used.
This research will be done by one person , but resources from other group members will be used where possible. These resources will consist of the knowledge and sources generated during the research of the other members. While not focusing on the same domain or question a lot of the base socio-ethical issues with the usage of AI are present over the multiple domains. There are no dependencies between any of the researches being done.

3 Issues of unfairness

Researching different examples of unfair outcomes within clinical decision support algorithms three types of issues appear to be the most common.
The first and most commonly talked about issue is the issue of bias towards a certain group of people or discriminatory effects and outcomes of decision support algorithms.
The second type of issue is that of clinical decision support algorithms disadvantageous to those who already have a disadvantage.
The third issue this paper will cover is that of unfairness caused by the way a clinical decision support algorithm is used.
In the following section an explanation and example of these three issues will be given after which the impact of that given issue on healthcare and people will be given.

The first issue of bias and discriminatory effects is one that many articles and research papers talk about. Not only about the presence of it but also the effect it could have on healthcare. And it is also one of the most used arguments against AI in healthcare. It is a problem not only present in clinical decision support systems but in many other algorithms even outside the field of healthcare. With one of the most common examples given, being negative bias towards inmates of colour in algorithms for the US justice system.

An example of such an issue within a clinical decision support system is for instance a commercial algorithm that guides health decisions[2]. A study into this algorithm showed that people who self-identified as black were generally assigned lower risk scores than equally sick white people[17]. The consequence this had was that black people were less likely to be referred to programs that aim to improve care for patients with complex medical needs. The study stated that fixing this issue would increase the percentage of black people receiving additional help from 17.7% to 46.5%([17].
The impact of biased and discriminatory decision support algorithms is seen on different layers. In the first place it gives an unfair advantage or disadvantage to a certain group of people which means not everyone is given equal level of care. Another consequence is that the bias and discrimination within these algorithms can lead to more bias and discrimination in the system as a whole.
The use of these algorithms brings the bias and discrimination within the the system leading to unequal access and resource allocation, discriminatory healthcare processes and biased clinical decision making. These inequalities then lead to discriminatory data taken from these systems[14]. This data is then again used to train, develop and research AI in healthcare, which then leads to more biased programs which then again are gonna be used in healthcare. This is a circle that repeats itself continuously worsening the discrimination and bias within the healthcare system.

The second type of issue commonly present in clinical decision support algorithms is algorithms that are disadvantageous to those who already have a disadvantage, may it be physically, economically or in another way.
Take for instance a predictive model for treatment recommendation. A model like this could make the choice to withhold certain treatment to a patient since the probability of benefiting is lower for this patient than another patient. This could be because of a genetic or chronic condition the first patient has, giving him a greater disadvantage within the algorithm.
A less probable but very plausible example lies within the abuse of algorithms for profit or cost reduction. Prediction and decision support algorithms give the tools for a hospital to identify vulnerable high-risk, high-cost patients and exclude them from care or act disadvantageous for the patients[8].
Another way this issue arises is the way certain patient data is collected. The smart watches and other wearable devices that have come onto the market in recent years have the ability to measure health data of users, such as heart rate, blood oxygen levels, physical activity and hours of sleep. Companies like Apple collect this data with the help of Apple Health. This collection of this data gives a very biased collection towards people that have access to and can afford wearable devices like this. When this data is then used for other algorithms and devices this bias is trained into these algorithms.

The impact of this type of issue is the added disadvantage towards the group that is already disadvantaged which can lead to another layer of disadvantages which then can lead to more disadvantage again. In addition to this this issue also gives potential to enlarge the gap between those who are and those who are not in poverty. When the data is becoming increasingly more biased towards those that already have an advantage, it becomes that much harder for those who are disadvantaged to get to the same level of care and being represented as those who are not.

The final issue we will cover in this section is regarding the unfairness caused by the way certain clinical algorithms are being used.
To give an example for this we can take the problem the US currently has with “No-Shows”. No shows is a term used for patients who fail to show up for a scheduled appointment. This is a major source of waste in the US healthcare[9] and the most effective way to combat this is overbooking[5]. AI can be used to assist with this issue by predicting which patients are least likely to not show up. This data can be predicted by the patient’s personal information such as clinical history, their patterns of healthcare use and the features of the appointment[9]. The downside and the issue we are talking about is when the patient that was predicted to “No-Show” and the overbooked patient both arrive. Now a slot planned for a single patient has to accommodate two patients leading to potential poorer quality of care[9]. So by using and trusting in the outcome of the AI used here efficiency goes up and wasted time goes down but also the quality of care goes down.

The impact that these types of issues have are more direct to the patients but influences the whole healthcare environment that uses it. It can lower quality of care in the example we gave, but can influence a lot more in other scenarios. There is not a single way this issue as a whole impacts healthcare, but it is the collective of these issues in which AI is being used in a way that can lead to unfair treatment of certain patients that is a problem and that can negatively impact the field of healthcare as a whole.

4 Causes of unfairness

The next section will go into what causes these aforementioned issues of unfairness of clinical decision support algorithms. The goal of this part is to create a better understanding in the underlying cause of these issues. By doing this we can better identify and pinpoint the problems that need to be solved or thought about. Since a lot of these issues require more than just a quick fix,the section after this will go into the socio-technical system design approaches and steps that exist to solve a lot of these problems. The causes mentioned will be the most common and impacting examples for clinical decision support systems. A lot of bias in systems has its own specific root of the problem and often many factors play a role when the issue of unfairness is found in a system, this next section will however try to get the base causes present in a lot of these cases.

The first cause and one that can be linked to almost all issues in some way is linked to bias in data. Almost all forms of algorithms and clinical decision support systems are trained and made with data. The problem is that this data is not always unbiased and by making an algorithm based on this data the bias within this data is brought into the algorithm. There are multiple ways bias can form or arise in data.
The first reason that data can be biased is because of the environment the data was taken from. If there is bias in the environment then the data taken from that environment will also be biased. A 2015 study from William J. Hall showed that among the healthcare providers from America most healthcare providers appeared to have implicit bias, mostly relating to racial and ethnic bias, impacting patient-provider decisions, treatment decisions, treatment adherence and patient health outcomes[12]. This implicit bias when taking information and data from this environment will be present in the data. The problem with this is that when a support system is designed or trained with this data this system will have this implicit bias and when used reinforce this bias making the bias only more present in the health care environment.
Another way that training data can become biased is by the way data is collected or chosen. Take for instance the deep learning convolutional neural network(CNN) that was created by a group of researchers mainly from Germany that was designed to recognize dermoscopic melanoma[11]. The CNN that was created was very successful and outperformed 58 dermatologists during testing. The CNN was created with help of data taken from the International Skin Imaging Collaboration, or ISIC, an open-source repository of skin images[1]. The problem was that the data of the ISIC had mostly images of white males making the CNN far less accurate on people of color.

One of the other problems here was also the way that the CNN was used. The CNN was made and designed by a mostly German group of researches and was also made with the German population in mind. The algorithm when then used in other places of the world for instance in America where the population is diverse leads to worse and sometimes even bias predictions[13].
This brings us to the next cause of unfairness and bias in algorithms and that is the difference in the way some algorithms are designed and how they are used. The previous example showed how when a clinical algorithm is used differently then the designers intended it can lead to a worse performance or bias.
This happens often when new breakthroughs are made that show great potential to do something good. As everyone is eager to use and apply it, the question of ”is this correct in our environment?” or ”does it work perfectly for everyone involved?” does not always get asked or answered.
This does not mean that algorithms should be made as a one size fits all. As David Leslie and colleagues explain in their paper [14] ”Every time a prediction model which has been

tailored to the members of a dominant group is applied in a "one-size-fits-all" manner to a disadvantaged group, the model might yield sub optimal results and be harmful for disadvantaged people.".
This is often linked with what is called the Portability Trap. As Andrew D. Selbst and his team explain in their paper on fairness and abstraction in sociotechnical systems [19]. The Portability trap states that within computer science it is often preferred to make programs and systems as abstract as possible so they can be used within different context. The same is true for machine learning algorithms grouping them in problems like clustering, prediction or classification. The issue that you create is that the system designed is very abstract and it misses the nuance and adaption needed for properly functioning in a very social setting.

Another cause of unfairness is the type of parameters or predictions being used to give the decision support. If we look back at the case of the commercial algorithm that guided health decisions by giving risk scores that gave overall lower risk scores to black people[17]. The problem found there that caused this discriminatory effect of these algorithms was with the way it predicted the risk score. This was done by predicting healthcare cost instead of illness of the patient, but unequal access to care means that we spend less money caring for black patients than for white patients. This means that while healthcare cost may look like an effective proxy for health it is one that has a lot of racial biases linked to it.

As David Leslie[14] tries to explain in his article, when features such as age, sex, ethnicity or socioeconomic status are integrated into models without careful consideration of potential confounders, those models risk identifying as biological, characteristics that have socioeconomic or environmental origins. As a result, structural racism might be integrated into the clinical decision support systems based on these models.
The lack of a variable/parameter can also cause unfairness in a clinical decision support system. For instance if data on skin colour are not collected together with pulse oximetry data, it is almost impossible for AI models to correct for the effect of skin tone on oximetry readings[14].

The last cause of unfairness created by clinical decision support algorithms that will be discussed will be unfairness caused by the use of a clinical decision support algorithm. This corresponds with for instance with the example of "No-Shows" given in the previous segment about the common issues.
If the algorithm was always 100% accurate then there would be no problem and it would be an amazing clinical support system that would increase efficiency. The problem is that it is not always 100% accurate and that by using this clinical support system you do increase efficiency but lower the quality of care for some people. By always aiming for the highest efficiency other aspects of healthcare like quality and equal care can suffer from this.
A different way for instance of using the aforementioned algorithm is by instead of double booking the slots of people that have a high chance of not attending their planned appointment, try to find a way to reach out to these people to work out a way to make sure they go to their appointment and get the care they need. As Sara G. Murray mentioned in her paper " Prior no-shows—a variable included in both the vendor's original model and our revised version—is likely to correlate with socioeconomic status, perhaps mediated by the inability to cover the costs of transportation or childcare, or the inability to take time away from work"[9]. This combined with the fact that most forms of AI are a "black box", meaning that you don't know what the contributing factors were for an algorithm making a decision. In this case this could mean that a patients slot is overbooked based on the fact that this patient is obese and because of this struggles with mobility. These patients do still need care, but have multiple of reasons for maybe not being able to attend their appointment.
By working these things out and communicating with the patients you increase the quality of care for these people and decrease the amount of no shows.

The way the previously discussed algorithm is used aligned with what is called the solutionism trap. Described in Andrew D. Selbst paper[19] as "Failure to recognize the possibility that the best solution to a problem may not involve technology". It describes how when there is a problem technology is often considered as one of the first solutions without looking at other possibilities, while other options work better in that scenario. In the case of "No-Shows" communications would be the better solution than relying on technology to predict who will not show up. This is not to say that the clinical support system can't help with this process but it shows how relying on a technical solution may not always be the most optimal solution.

5 Socio-technical design

This section will introduce different aspects found in socio-technical design methodologies. There are multiple methodologies with their own rules and guidelines on how to design and manage socio-technical systems. Their are a lot of common aspects and focuses between the different methodologies. This section will try and explain these common aspects as well as some stand alone methodologies. Small differences between methodologies will be highlighted and evaluated. A problem with a lot of these rules and guidelines is that they are very broad and are hard to apply to specific scenarios or systems. This is why in the next section these design approaches will be adapted to the specifics of designing for clinical decision support algorithms.

The first common idea in a lot of the socio-technical design methodologies is understanding the problem. This often consists of many steps and it considers many of the different factors tied to the problem. The meaning of understanding the problem while often in it's base form being the same differs a lot from one methodology to the other. By going along a standard process this paper will try to explain the differences and similarities.

The first step in many of the different methodologies involves a form of stakeholder analysis. To really understand and solve a problem it is important to know the different stakeholders. When talking about stakeholders most methodologies try to involve all people that are in some way connected to the problem or the sociotechnical system that is being designed. These are people that will work with your system, people that are influenced by the system or any other group of people that is in some way connected or affected by either the system or the problem. How we analyze these stakeholders differs per methodology. In the case of the Soft Systems Methodology [7] it takes into account the roles, responsibilities and concerns of the stakeholders. While in the case of designing for value [20] it analyzes the values and norms of the different stakeholders. While yet other methodologies such as contextual design[6] and human centred design [4] focus more on the people using the system or customers of the system in some way. Asking how they are affected by the problem without the system, how they will interact with the system, in what kind of environment they work or interact with the problem and system and what their wants and needs are.

Following the step of understanding the problem and analysing the stakeholders we can look at the guidelines and methodologies surrounding the technology part of the socio-technical design approach. At first we have the Cognitive Work Analysis [18] in this methodology it focuses on a socio-technical system in the terms of what it can do instead of a more normative and usual approach of asking how work should be done or how work is done. By analysing the technology in this way, the method tries to understand what the capabilities of the technology are. In this way avoiding using a technology in a place where it does not fit or for a purpose it can't serve. With this also comes a analysis that is done in many of the methodologies, in which the designers try to identify tasks that have to be allocated to machines and tasks that have to be performed by humans, both individually and together. By identifying these tasks a system can be designed to better fit humans and technology working together.

The next step of the process will focus on the actual designing and making aspect of socio-technical design methodologies. Within the designing step there are a lot of nuances tied to the system that is being designed. Nonetheless there are socio-technical design methodologies that give a basis for developing a system. In the paper of Selbst and his team they explain the different traps of designing a socio-technical system and how to work around those traps[19]. One of the traps he gives is the Formalism Trap. The Formalism trap is explained as the problem of implementing fairness as a mathematical problem. The problem with this is that fairness is something very contextual and changing and is hard to state as a mathematical problem, as we would like to do with machine learning. This would mean that maybe fairness is just not implementable in software and that we should solve this problem otherwise, but in Selbst's paper he suggests something called the Social Construction of Technology program(SCOT)[14]. How Selbst explains this is "The key elements of the SCOT framework are a period of interpretive flexibility experienced by relevant social groups, followed by stabilization, and eventually closure."[15]. What this means is that when new technologies are introduced, many different versions will be used and each of these different versions will adapt to the social setting that they will be used in. Adapting and changing until the relevant social group redeems it fair/ considers the problem solved. By adapting versions to the local problems and idea of fairness it is much better feasible to solve the problem of fairness within a system. A lot of the other methodologies align with the way of thinking described here, as in that the designed system should be adapted to the stakeholders and the environment it will be used in. Where again the stakeholder the methodology focuses on decides the design of the system.

Another important step in designing a socio-technical system is designing for the influence the system will have on the environment and the people around it. Within the different methodologies this is often referred to as the ripple effect or the ripple effect trap [19].The ripple effect trap deals with the consequences of introducing new technology into a system.Introducing a new technology into a system can have an effect of a lot of different factors, mainly the people working and interacting within this system. To solve this seems almost impossible since it would require knowing/predicting what the influence is of introducing the technology into a new environment. Fortunately this is easier than it sounds. A lot of the common changes a piece of technology can create are well documented and understood. By being aware of these changes and designing with preventing the negative ripple effect in mind it becomes a lot easier to create a system that does not cause big ripple effect when being introduced into a system.

As seen there are a lot of common ideas within the different socio-technical design methodologies. With the main idea of creating a better understanding of the problem, the environment and the different parts linked to the sociotechnical system. Which is then used to design with this better understanding. Designing for the different important aspects involved in a sociotechnical system.

6 Adapting for Healthcare

All the design approaches mentioned before form a good basis for a step by step guideline for improving and designing socio-technical systems, but what it lacks is the more in depth adaptation needed for designing clinical decision support systems for fairness. This is what this next section will focus on. The design approaches will be adapted to the problems and causes mentioned earlier in this paper adding and changing where necessary to fit with the intricacies involved with clinical decision support systems.

The designing of a clinical decision support system should start with either one of two ways. Either there is a problem or there is a new technology that can be used as a clinical decision support system to solve a problem. In both cases the processes are very similar, with the slight difference being that in the case of a new technology a lot of the designing process is changed into an adapting process.

In the case that there is a problem, the process of designing should start off by analysing and understanding the problem. There are a few factors of the problem that need to be understood before starting to design a clinical decision support system to solve it. In the first place it needs to be clear who the stakeholders are. In the case of clinical decision support systems these are mostly the following:
Healthcare workers that encounter this problem and who will have to work with the designed system.
Patients that will either interact with your system or are in some way influenced by the system.
The hospitals or clinics that choose to use this system.
Other stakeholders might be insurance company's, the government or something else linked to or dependent on the problem or linked to the specific clinical decision support system. For each of these stakeholders it needs to be clear what their roles are within this problem and within the future system.
Will they need to make choices based on the outcome of a system or are they just influenced by the system. What are their responsibilities within the scope of this problem.
Next you want to know how these stakeholders deal with the problem currently. By understanding what their solutions or personal problems within this problem are there is a better vision on what the clinical decision support system will have to be able to do. After analysing the different aspects of the problem and analysing the stakeholders you can start to look towards a solution. After analysing the problem you will also have a good answer on the question do we need a technological solution here or is there another maybe more fair solution.

The next step in the process is actually designing and making the system. Within this step there are a few steps to take and a few pitfalls to avoid to ensure fairness.
First again with the knowledge of who the stakeholders are within the designed system. Design for the needs, wants, concerns and other important features of the stakeholders and involve them in the process early on. When deciding what parameters and proxy's to use when making a prediction algorithm think of the possible bias within these parameters. Do these parameters favor one group of people over another and if so is it replaceable for another parameter or should there be a correction in the training data. If the bias can not be fixed maybe there should be two different support systems to give equal chances. Don't try to force a one size fits all but adopt and use smaller versions adapted to certain groups or environments if this means better accuracy for all people involved. When choosing how to collect data to use for designing and making the support system again consider all the stakeholders. Does the data represent everyone equally and does it fit the specific problem and stakeholder the system is designed for.

At last after testing the support system, making sure the test are re presentable for the environment in which the system is going to be used, there are still some steps to applying the support system to the environment. In this step it is necessary to ensure that again all involved stakeholders know what the limits of the system are. When using the systems the people that are going to be working with the system should know what the actual outcomes of the systems mean and how they should be used.

7 Discussion & Conclusion

This section will try to explain and discuss some of the still unanswered question, discussion points and the future work into this subject. It also reflects on some of the decisions made for this paper and discusses why they were made.

One of the discussion points important to this paper is the role of computer scientist in all this. As a computer scientist fairness and socio-ethics are often not at the front of the priority list and are often overshadowed by factors such as speed, accuracy and cost efficiency. While often for computer scientist socio-ethics and fairness might not be the first thing they think about it should be something that is central in the design process. As designers, makers and testers of the systems computer scientist have the possibility of designing for the socio-ethical implication and making the system as fair as possible. However as stated in this paper this is not feasible without communication with the different groups of stakeholders connected to the the system. To ensure fairness all stakeholders should be represented and work together. Trying to understand the steps that need to be taken to design a fair decision support system for everyone. As a fair decision support system even has the possibility to reduce the already existing bias within healthcare by designing support systems that do not re enforce the bias that is already in the system, but instead aim for equal and fair care for everybody.

Import as well in the whole process from understanding, designing, adapting and applying is to leave a transparent path in the work done. This will ensure that mistakes can more easily be spotted and that different decision from the system but also around the system can be explained more easily. This also reduces/prevents people with bad intent to purposefully enter bias or use or design the system in a way that would be unfair.

Another point of discussion are the examples used within this paper. Most of them where examples of the USA healthcare system. The choice for this wasn't because of preference or familiarity from the author. These examples were chosen because of the most available reliable information that could be found online. This does not mean that the conclusion written is not applicable to designing clinical decision support systems for other countries. This does mean however that there may be different bias or ripple effects to look out for and different rules to work around. The main take away is that every project and system is different and adapting and designing for that specific project and environment is the most important.

At last we have the limitation of this paper. The information provided within this paper is based on a literature study from papers taken from the different fields. Theses being the computer science, healthcare and social study's fields. While many of the separate parts of the methodology for designing fair clinical decision support systems have been tested, the combined and adapted methodology as a whole has not been proven by any empirical research.
Future research will need to be done, applying the ideas from this paper to actual design of clinical decision support systems. Refining and expanding upon the steps given in this paper to create better and more reliable resources for creating fair clinical decision support systems.

So to conclude within this paper an overview has been given on the problem of fairness within clinical decision support systems, explaining the effect these problems have on people and the healthcare system. After which the different underlying causes for these problems have been explained. Following this a overview has been provided on the different socio-technical design methodologies that exist, indicating the similarities and differences between the different methodologies. At last a separate methodology for designing fair clinical decision support systems is suggested and explained, combining and adapting the previous methodologies according to the problems and causes mentioned within this paper.

8 Responsible research

Within this research and paper I the author have tried to stay as objective as possible. While searching and reviewing different papers many scientifically valid papers were taken into account and none were excluded for opposing or contradicting theories or arguments. All statistics and examples explained in this paper came from scientific papers or research done by trustworthy sources. While the last part is a combined composition of all the sub sections before it some subjective influence was put in by the author. The subjective part being the conclusion from the previous sub sections on how to best ensure fairness into the design process of clinical decision algorithms. To add to that this whole paper and research relies on the subjective view of what it means for somethign to be fair. While the choice has been made to keep an as objective view on fairness as possible, others may draw different conclusions or have a different opinion of fairness. All referencing for forming ones own opinion and conclusion are given in the reference section and throughout the relevant parts of the text.

References

[1] Isic archive. (n.d.). isic. retrieved june 16, 2021, from https://www.isic-archive.com/!/topwithheader/widecontenttop/main.

[2] What's medicare? (n.d.). medicare. retrieved june 20, 2021, from https://www.medicare.gov/what-medicare-covers/your-medicare-coverage-choices/whats-medicare.

[3] Evgeni Aizenberg and Jeroen van den Hoven. Designing for human rights in ai. *Big Data & Society*, 7(2):2053951720949566, July 2020. Publisher: SAGE Publications Ltd.

[4] Gordon Baxter and Ian Sommerville. Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1):4–17, 08 2010.

[5] Murr M. Chermak D. Woodall J. Pignone M. Sandler R. S. Denton-B. T. Berg, B. P. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Medical decision making : an international journal of the Society for Medical Decision Making*, 33(8):976–958, 2013.

[6] Hugh Beyer and Karen Holtzblatt. Contextual design. *Interactions*, 6(1):32–42, January 1999.

[7] Peter Checkland and J Poulter. *Learning for Action: A Short Definitive Account of Soft Systems Methodology and its Use, for Practitioners, Teachers and Students*. John Wiley and Sons Ltd, 2006.

[8] I. Glenn Cohen, Ruben Amarasingham, Anand Shah, Bin Xie, and Bernard Lo. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Affairs*, 33(7):1139–1147, 2014. PMID: 25006139.

[9] Sara G. Murray Robert M. Wachter Russell J. Cucina. Discrimination by artificial intelligence in a commercial electronic health record—a case study. *Health Affairs Blog*, 2020.

[10] Chiara Garattini, Jade Raffle, Dewi N. Aisyah, Felicity Sartain, and Zisis Kozlakidis. Big data analytics, infectious diseases and associated ethical impacts. *Philosophy & Technology*, 32(1):69–85, March 2019.

[11] Schneiderbauer R Toberer F Buhl T Blum A Kalloo A Hassen ABH-Thomas L Enk A Uhlmann L; Reader study level-I Haenssle HA, Fink C and Arenbergerova M Bakos R Baltzer A Bertlich I Blum A Bokor-Billmann T Bowling J Braghiroli N Braun R Buder-Bakhaya K Buhl T Cabo H Cabrijan L Cevic N Classen A Deltgen D Fink C Georgieva I Hakim-Meibodi LE Hanner S Hartmann F Hartmann J Haus G Hoxha E Karls R Koga H Kreusch J Lallas A Majenka P Marghoob A Massone C Mekokishvili L Mestel D Meyer V Neuberger A Nielsen K Oliviero M Pampena R Paoli J Pawlik E Rao B Rendon A Russo T Sadek A Samhaber K Schneiderbauer R Schweizer A Toberer F Trennheuser L Vlahova L Wald A Winkler J Wölbing P Zalaudek I. level II Groups, Alt C. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol.*, 29(8):1836–1842, 2018.

[12] Chapman M. V. Lee K. M. Merino Y. M. Thomas T. W. Payne-B. K. Eng E. Day S. H. Coyne-Beasley T. Hall, W. J. "implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: A systematic review". 2015.

[13] Angela Lashbrook. "ai-driven dermatology could leave dark-skinned patients behind". 2018.

[14] David Leslie, Anjali Mazumder, Aidan Peppin, Maria K Wolters, and Alexa Hagerty. Does "ai" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*, 372, 2021.

[15] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):2053951716679679, December 2016. Publisher: SAGE Publications Ltd.

[16] Jessica Morley, Caio C. V. Machado, Christopher Burr, Josh Cowls, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. The ethics of ai in health care: A mapping review. *Social Science & Medicine*, 260:113172, September 2020.

[17] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

[18] Jens Rasmussen, Annelise Mark Pejtersen, and L. P. Goodstein. *Cognitive Systems Engineering*. John Wiley Sons, Inc., USA, 1994.

[19] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 59–68, New York, NY, USA, 2019. Association for Computing Machinery.

[20] MJ van den Hoven, PE Vermaas, and IR van de Poel. *Design for values: An introduction*, pages 1–7. Springer, 2015. Harvest.