A Review of Algorithmic Bias in the American Healthcare System

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Katherine M. Taylor

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

Introduction and Methods

The COMPAS Algorithm Controversy

The COMPAS recidivism algorithm sounded like an excellent idea. The sentencing of someone who has committed a crime can easily be affected by human bias. Humans can be racist, classist, and sexist without even realizing it. The natural, obvious solution was to allow an algorithm to make this decision. Surely an algorithm would be less biased than a human. After all, an algorithm makes decisions based on real data without any input from feelings, upbringing, or anecdotal experience. And yet racial bias in the COMPAS algorithm started to become apparent. Prisoners of color were more likely to be deemed dangerous, even when they did not commit a second crime. White prisoners who eventually did re-offend were marked as low risk (Martin, 2019; Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021; Yates, Gulati, & Weiss, 2013). Clearly, algorithms can be just as biased as humans.

As the use of algorithms in everyday life has increased in recent years, the necessity of recognizing bias in machine learning algorithms has become progressively more important. Many types of bias can infiltrate a machine learning algorithm, usually as a result of poor feature selection, bad training data or lack of transparency in the algorithm creation process (Martin, 2019). The COMPAS recidivism algorithm was marked both by poor feature selection and a lack of transparency. Among other characteristics, an inmate's neighborhood and family crime history were considered in the decision (Martin, 2019; Mehrabi et al., 2021; Yates et al., 2013). This poor feature selection ended in incorrect predictions of recidivism across races. Furthermore, the judges that employed the algorithm had almost no knowledge of how the algorithm worked "under the hood." This lack of transparency led to judges being given a number spit out by the algorithm with little to no information on how to weigh this number with the other, human-determined factors (Martin, 2019; Mehrabi et al., 2021; Yates et al., 2013). There have been multiple instances of algorithms that failed due to bad training data; natural language processing reflects gender bias and object recognition algorithms correctly identify images from Western culture at a higher rate than images from Eastern cultures (de Vries, Misra, Wang, & van der Maaten, 2019; Leavy, 2018).

Bias in machine learning algorithms has the potential for extremely deadly consequences. Human bias, both implicit and explicit, already plagues the healthcare field. Doctors are more likely to downplay pain reported by women than pain reported by men (Colameco, Becker, & Simpson, 1983). Racial bias is particularly rampant in healthcare; for example, Black women have disproportionately high rates of maternal death when compared with White women (Hall et al., 2015; Rosenberg, Geller, Studee, & Cox, 2006). This existing bias can easily leach into data collection for algorithms if healthcare professionals and developers are not careful (Starke, De Clercq, & Elger, 2021). The presence of a sociotechnical framework might assist these stakeholders in their search for a fair data set and ensuing algorithm.

The Responsible Innovation Framework

The Responsible Innovation framework, developed by Jack Stilgoe et al. (2013), seeks to give a detailed plan for developing a technology with as few negative consequences as possible. Stilgoe first acknowledges that most innovators seek to develop their products responsibly without knowingly inserting bias into their algorithms. However, unintended consequences that negatively affect one or more populations are a common occurrence in technology development. In order to recognize and prevent unintended consequences such as bias, developers should consider a framework with four critical dimensions: anticipation, reflectiveness, deliberation, and responsiveness (Stilgoe, Owen, & Macnaghten, 2013).

Anticipation asks the question “what if?” It seeks to determine the possible unintended consequences by thoroughly examining the technology. Reflectiveness is an extension of anticipation and seeks to discuss what is known, such as the purposes and motivations of the technology, as well as what is unknown, such as areas of ignorance and questions. Deliberation entails the opening of this conversation about possible biases to stakeholders. This allows for a collaborative environment that acknowledges a variety of perspectives. Responsiveness is a secondary, iterative measure that reflects on past progress in order to direct the future progression of the technology (Stilgoe et al., 2013). With respect to bias, anticipation and reflection entail the examination of the dataset, methods of data collection, and the algorithm itself for potential sources of bias, with an emphasis on analyzing ways the algorithm may behave unexpectedly. Deliberation with stakeholders may open developers up to their own implicit biases. Finally, biases in a machine learning algorithm should be investigated after feature extraction, training, validation, and testing as a form of responsiveness.

In contrast to responsible innovation is irresponsible innovation, stemming from at least one of four practices ranging from thoughtless to intentionally deceptive and resulting in problematic technologies such as biased algorithms. A technology push occurs when one stakeholder pushes a feature in the algorithm that leads to dissent from others. The neglecting of fundamental ethical principles during early stages of development results when researchers decide that considering ethics is less important than rolling out the first prototype. A policy pull occurs when there is a strong social desire to produce the technology as fast as possible. Finally, a lack of precautionary measures and technological foresight occurs as a result of uninformed researchers or a willful ignorance of the possibility of ethical issues within an algorithm (von Schomberg, 2013).

When moving from the theoretical to the practical, the responsible innovation framework offers several concrete examples of both proactive and irresponsible behavior during the development of a technology such as a machine learning algorithm. For example, the ethics of an algorithm should be a design factor rather than a constraint (von Schomberg, 2013). The goal should move from making an algorithm that won't cause any trouble to making an algorithm that serves all targeted stakeholders equitably. In contrast, the tendency to produce technology as quickly as possible due to demands from stakeholders can easily result in unintended consequences (von Schomberg, 2013). Further research will be conducted on specific action items necessary for bias mitigation within machine learning algorithms, specifically within the context of healthcare.

Research Methods

The responsible innovation framework leads me to ask: How can sources of bias be mitigated in machine learning algorithms, especially algorithms that involve the healthcare sector? This question will be answered through a comprehensive literature review. The review began with a discussion of algorithms that failed due to bias. The aspect of the responsible innovation framework that is most often ignored was determined from this section of the review. Next, articles that describe a machine learning algorithm associated with at least one bias mitigation technique were reviewed; these mitigation techniques were also analyzed within the context of the responsible innovation framework. Results from these two separate reviews were synthesized into a final report detailing the current state of bias mitigation in terms of the responsible innovation framework.

Results and Analysis

Algorithms Failing Due to Bias

ImpactPro

The ImpactPro algorithm was created in an attempt to identify high-risk patients based on a number of quantifiable factors, with the intention of enrolling patients with the highest risk in a specialized health management program. The algorithm assigned a risk score to each patient based on their electronic medical record, which included past hospital stays, chronic conditions, and insurance information, among other things. If the risk score was higher than the 97th percentile, the patient was automatically referred to a health management program. If this risk score was higher than the 55th percentile, the patient was referred back to their primary care doctor to determine whether an enrollment was necessary (Obermeyer, Powers, Vogeli, & Mullainathan, 2019). The intention behind the development of this algorithm was multi-fold: healthcare providers could identify patients in most need of extra care management while simultaneously saving costs (Sargent, 2021).

A study testing the effectiveness of ImpactPro eventually discovered several disturbing things about the outputs of the algorithm. Patients who identified as Black represented 17.7% of patients enrolled in the program, but should have made up 46.5% of the enrollees (Obermeyer et al., 2019). Within a given risk score percentile, Black patients had higher hypertension, more severe diabetes, increased LDL (bad cholesterol), higher creatinine levels (a mark of renal failure), and lower hematocrit (a mark of anemia) when compared with White patients (Obermeyer et al., 2019). Overall, they were significantly less healthy than White patients and had a greater number of chronic conditions. The percentage of Black patients enrolled in the program after a consultation with their primary care provider was higher than the percentage automatically enrolled by ImpactPro, but still did not reflect the percentage of Black patients needing extra care management (Sargent, 2021).

Further research into the inner workings of ImpactPro yielded the explanation behind this bias. One of the features that determined the risk score was the healthcare costs the patient had accumulated over time (Obermeyer et al., 2019; Sargent, 2021). Past healthcare costs were intended to be used as a prediction for future healthcare costs; higher predicted healthcare costs translated to higher risk. Overall, Black patients accumulated lower healthcare costs when compared with White patients with the same health problems, resulting in the observed lower risk scores (Obermeyer et al., 2019; Sargent, 2021). This cost phenomenon can be explained by the fact that Black patients have a higher poverty rate than White patients, which can reduce access to care. This lack of access was supported by data that showed that Black patients attributed most of their healthcare costs to emergency procedures, while white patients attributed most of their healthcare costs to outpatient and elective procedures. Furthermore, Black patients have a distrust of the healthcare system due to past adverse experiences such as the Tuskegee study, which can discourage Black patients from seeking care (Obermeyer et al., 2019).

This case study fortunately has a happy ending. The company that developed ImpactPro realized their mistake and produced a new algorithm that was solely based on a quantification of health data (Price, 2019). This algorithm ended up being less biased than the human primary care managers and reduced racial disparity by 84% (Price, 2019; Sargent, 2021).

Skin Cancer Prediction

One promising development in the healthcare field is the use of machine learning to classify an image of an abnormal skin growth as benign or malignant. These classification algorithms utilize large, publicly available skin image datasets to train. The accuracy of these algorithms is therefore largely dependent on the quality of the dataset. Multiple studies have detailed either theoretical or observed issues with datasets that can inject bias into an algorithm.

It takes a certain level of difficulty to correctly classify a given skin image, and many skin image datasets are primarily comprised of low and medium difficulty datasets which are generally used as teaching material. This results in classification algorithms finding images with a high level of difficulty unusually difficult, which is the exact opposite of the desired phenomenon (Bissoto, Fornaciali, Valle, & Avila, 2019). More importantly, these skin image datasets are almost always lacking in diversity of skin tone. Lack of diversity in skin tone means that the algorithm overfits to that particular skin tone and translates poorly to other skin tones (Wen et al., 2022). Furthermore, the datasets are very Western-centric. Studies have found that the majority of datasets are from Europe, North America, or Oceania, with no datasets from Africa; each dataset typically only has images from one country as well. Within datasets that report ethnicity, there are typically only a few out of thousands of images that have the Fitzgerald skin type V or VI, which corresponds to darker skin (Guo, Lee, Kassamali, Mita, & Nambudiri, 2021; Wen et al., 2022). This phenomenon, coupled with the already existing dearth of skin of color in dermatological teaching materials, points to a necessity for greater inclusivity in skin image datasets (Guo et al., 2021).

Kidney Failure Risk Equations

Kidney failure risk equations (KFREs) are used clinically to determine the risk of end-stage renal disease (ESRD). One popular equation developed by Navdeep Tangri utilizes 8 variables including age, sex, estimated glomerular filtration rate (eGFR), and urine albumin to creatinine ratio (ACR), to produce percentages reflecting the likelihood of ESRD in two and five years. This percentage determines next steps in the treatment plan for the patient. The most obvious source of bias is the necessity for the eGFR and ACR to even apply the equation; these two statistics require a blood test and urine test, respectively. Younger patients are less likely to be tested for eGFR and ACR as they are seemingly healthier (Williams & Razavian, 2019). Furthermore, lab work can present a monetary burden to patients of lower socioeconomic status. Researchers called into question the accuracy of this equation across genders, ages, and races. Though the model performed well when used on children (Winnicki et al., 2018), there were pitfalls in other categories.

The Tangri equation overestimates the likelihood of ESRD for patients that were over 80 years old. This is mostly due to the increased likelihood of death from unrelated causes within the next two or five years (Hundemer et al., 2021). However, this overestimation was not true for Black patients over the age of 80. This is due to the documented faster progression to ESRD that occurs in patients of African descent (Ahmed et al., 2021; Grams et al., 2015). Furthermore, some KFREs use a race multiplier that multiplies eGFR by 1.159 for Black patients, which stems from the belief that Black patients naturally have higher muscle mass than White patients (Ahmed et al., 2021). Since a lower eGFR is indicative of a higher risk, this meant that Black patients were given an estimation of risk that was far lower than the actual risk. When removing this race multiplier, one in three Black patients was classified to a more severe stage of kidney disease (Ahmed et al., 2021). The lack of understanding regarding Black patients and kidney failure likely stems from the lack of inclusion in research studies. Two studies validating the Tangri equation were 73% and 92% Caucasian, which is not indicative of the general population (Hundemer et al., 2021; Peeters et al., 2013). Overall, the Tangri equation is very effective at predicting ESRD for old, white men, but could be improved or replaced to be more effective when predicting for more marginalized populations.

Reasons for Algorithm Failure

As mentioned previously, bias in a machine learning algorithm is usually due to one of three things: poor feature selection, bad training data or lack of transparency in the algorithm creation process (Martin, 2019). The algorithms studied above that failed due to bias can be classified into these three categories with ease: ImpactPro suffered from poor feature selection as well as a lack of transparency, resulting in primary care providers not knowing that the algorithm was racially biased. The skin cancer detection algorithms suffer from a lack of diversity in the training data. The Tangri equation also suffers from poor feature selection. While that is easy to determine, it is harder to pinpoint exactly why these algorithms ended up with poor features, bad training data, or lack of transparency. This will be examined in the context of the responsible innovation framework.

ImpactPro mostly suffered from a lack of precautionary measures. A few simple graphs before implementation of the algorithm would have been sufficient to show that Black patients were being accepted into the care program at a much lower rate than they should have been accepted giving the prevalence of chronic conditions within their race. Skin cancer prediction algorithms mostly suffered from the neglecting of fundamental ethical principles, which is reflected in the available datasets. Had the issue of skin color been raised beforehand, it is likely that datasets would span multiple countries and continents to provide a more expansive view of skin throughout the world. Finally, the Tangri equation was also marred by a consistent neglect of ethical principles. Though Tangri did validate his model on different demographics, such as different ages and countries of origin, most of the patients from the study were White. Furthermore, the use of a race multiplier to falsely increase eGFR levels in Black patients is particularly offensive. All three case studies suffered from the policy pull phenomenon to some extent. There is a pervasive theory circling the globe that technology is more accurate than humans, and there is much pressure to achieve diagnostic accuracy within the healthcare sector.

Having identified lack of precautionary measures and neglect of ethical principles as some of the most prevalent reasons for algorithm failure, I will now turn to the mitigation techniques that are currently in use in healthcare algorithms in the context of the responsible innovation framework.

Algorithms Succeeding due to Mitigation Techniques

Improved Tangri Model

One study tried to improve the reachability of KFREs by creating a machine learning model that solely utilized information from electronic medical records. They found that 95% of patients who progressed to renal failure did not have the required variables for the Tangri equation, meaning that no risk could be predicted (Williams & Razavian, 2019). The model was trained on an expansive, diverse dataset of 1.6 million people who had received care from NYU Langone Medical Center (Williams & Razavian, 2019). Due to the diversity of the training data, the model not only outperformed the Tangri equation in accuracy for female, Black, Asian, and younger cohorts, but had more comparable accuracy scores across all cohorts (Williams & Razavian, 2019). This study represented a positive step towards a more equitable KFRE and showed the usefulness of a larger and more diverse training set.

Fahrenbach Hospital Stay Algorithm

Another study led by John Fahrenbach halted its progress when it determined that the developing algorithm was racially biased against Black patients. Initially, the researchers hoped to create a model that would predict which patients were most likely to be discharged from the hospital in 48 hours. These patients would then receive extra care management including discharge instructions and administrative assistance (Conkling & Marsh, 2022). When figuring out which labels were best suited for predicting hospital stay length, researchers determined that the patient’s zip code was particularly helpful in prediction (Conkling & Marsh, 2022; Strickland, 2019). Rather than going ahead with the creation of the model, this label choice was scrutinized. It was determined that the zip code reflected the city’s racial and socioeconomic demographics: affluent and typically White patients had shorter hospital stays than less affluent and typically Black patients (Conkling & Marsh, 2022). If this model had been implemented, resources would have shunted towards the more affluent patients, and less affluent patients may have been deprived of necessary care.

After realizing their mistake, the researchers reached out to a diversity and inclusion team to determine the best next steps. They determined a basic protocol that includes checkpoints throughout the development process to ensure that the algorithm being developed is not biased (Conkling & Marsh, 2022). Although this is not something that has been implemented on a national or global scale, it represents a good start for mitigating bias specifically within algorithms intended to improve healthcare.

Current Mitigation Techniques

The improved Tangri model and Fahrenbach hospital stay algorithm both used bias mitigation techniques successfully to make algorithms that were fairer than their predecessors. The improved KFRE equation utilized a larger, more diverse training data set and tried a different learning approach to solve this problem. This addresses the reflective and responsive aspects of responsible innovation: the researchers realized that the original equation was racially biased and iterated through another approach to improve the overall equation. The fallout from Fahrenbach’s hospital stay algorithm resulted in a protocol that resonates with responsible innovation. Examining the dataset with bias in mind is critical for the anticipatory dimension of the framework. Establishing checkpoints within the development process to ensure the model is still acting fairly is paramount to the reflective dimension. Engaging with other people such as ethics professionals, providers, and patients during the process is a perfect illustration of the deliberative dimension, and monitoring the system after development sets developers up to be responsive in the case of bias. Overall, while there are certainly examples of algorithms that failed due to bias, the developing awareness for the necessity of responsible innovation when creating an algorithm is promising for the future.

Discussion

The impact of a biased algorithm in healthcare can be dangerous to human life. As mentioned previously, algorithms are generally viewed as an unbiased means to bring equitable healthcare to all patients. With the rising concern for implicit human bias within healthcare, machine learning is becoming more popular. Unfortunately, studies that uncover algorithmic bias that is worse than human bias in some cases have emerged.

If these biased algorithms continue to be used, there is the strong possibility that communities already marginalized in the healthcare sector may become even more marginalized. An even more frightening thought is that healthcare providers may sensibly believe that the decisions made by the algorithm are fair and less biased than decisions that they would make. As medical schools become more cognizant of bias present in healthcare and work to inform future providers of their own implicit biases, biased algorithms may be a step backward. Communities marginalized in healthcare already have a general mistrust of the healthcare system, and the presence of biased algorithms will only increase this mistrust, which can lead to avoidance of necessary care.

The protocol implemented by Fahrenbach is a first step in preventing irresponsible innovation because it addresses all four tenets of responsible innovation. A protocol that asks developers to approach a problem with awareness about bias, reflect on previous failures, ask for guidance from other stakeholders (including ethical professionals), and iterate through new models when necessary would go a long way in preventing bias if this was implemented globally for every machine learning project. In addition to a standardized protocol for healthcare algorithms, educating developers, healthcare providers, and the general public about potential bias in healthcare algorithms will provide all stakeholders with information about algorithms and how to prevent them from being biased. Overall, the emerging awareness about bias in algorithms along with protocols that follow the responsible innovation framework will hopefully lead to a decrease in bias with the healthcare sector as well as a feeling of increased safety for those most marginalized by healthcare professionals.

References

Ahmed, S., Nutt, C. T., Eneanya, N. D., Reese, P. P., Sivashanker, K., Morse, M., … Mendu, M. L. (2021). Examining the Potential Impact of Race Multiplier Utilization in Estimated Glomerular Filtration Rate Calculation on African-American Care Outcomes. *Journal of General Internal Medicine*, *36*(2), 464–471. https://doi.org/10.1007/s11606-020-06280-5

Bissoto, A., Fornaciali, M., Valle, E., & Avila, S. (2019). (De) Constructing Bias on Skin Lesion Datasets. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2766–2774. Long Beach, CA, USA: IEEE. https://doi.org/10.1109/CVPRW.2019.00335

Colameco, S., Becker, L. A., & Simpson, M. (1983). Sex Bias in the Assessment of Patient Complaints. *THE JOURNAL OF FAMILY PRACTICE*, *16*(6), 5.

Conkling, B., & Marsh, C. (2022). Identifying Bias in Hospital Length of Stay Algorithm. Retrieved February 7, 2022, from Booz Allen Hamilton website: https://www.boozallen.com/c/insight/blog/identifying-bias-in-hospital-length-of-stay-algorithm.html

de Vries, T., Misra, I., Wang, C., & van der Maaten, L. (2019). *Does Object Recognition Work for Everyone?* 52–59. Retrieved from https://openaccess.thecvf.com/content_CVPRW_2019/html/cv4gc/de_Vries_Does_Object_Recognition_Work_for_Everyone_CVPRW_2019_paper.html

Grams, M. E., Li, L., Greene, T. H., Tin, A., Sang, Y., Kao, W. H. L., … Appel, L. J. (2015). Estimating Time to ESRD Using Kidney Failure Risk Equations: Results From the African American Study of Kidney Disease and Hypertension (AASK). *American Journal of Kidney Diseases*, *65*(3), 394–402. https://doi.org/10.1053/j.ajkd.2014.07.026

Guo, L. N., Lee, M. S., Kassamali, B., Mita, C., & Nambudiri, V. E. (2021). Bias in, bias out: Underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection—A scoping review. *Journal of the American Academy of Dermatology*. https://doi.org/10.1016/j.jaad.2021.06.884

Hall, W. J., Chapman, M. V., Lee, K. M., Merino, Y. M., Thomas, T. W., Payne, B. K., … Coyne-Beasley, T. (2015). Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *American Journal of Public Health*, *105*(12), e60–e76. https://doi.org/10.2105/AJPH.2015.302903

Hundemer, G. L., Tangri, N., Sood, M. M., Clark, E. G., Canney, M., Edwards, C., … Akbari, A. (2021). The Effect of Age on Performance of the Kidney Failure Risk Equation in Advanced CKD. *Kidney International Reports*, *6*(12), 2993–3001. https://doi.org/10.1016/j.ekir.2021.09.006

Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. Gothenburg Sweden: ACM. https://doi.org/10.1145/3195570.3195580

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Peeters, M. J., van Zuilen, A. D., van den Brand, J. A. J. G., Bots, M. L., Blankestijn, P. J., Wetzels, J. F. M., … van Zuilen, A. D. (2013). Validation of the kidney failure risk equation in European CKD patients. *Nephrology Dialysis Transplantation*, *28*(7), 1773–1779. https://doi.org/10.1093/ndt/gft063

Price, M. (2019, October 24). Hospital 'risk scores' prioritize white patients. Retrieved February 6, 2022, from Science website: https://www.science.org/content/article/hospital-risk-scores-prioritize-white-patients

Rosenberg, D., Geller, S. E., Studee, L., & Cox, S. M. (2006). Disparities in Mortality Among High Risk Pregnant Women in Illinois: A Population Based Study. *Annals of Epidemiology*, *16*(1), 26–32. https://doi.org/10.1016/j.annepidem.2005.04.007

Sargent, S. L. (2021). AI Bias in Healthcare: Using ImpactPro as a Case Study for Healthcare Practitioners' Duties to Engage in Anti-Bias Measures. *Canadian Journal of Bioethics*, *4*(1), 112–116. https://doi.org/10.7202/1077639ar

Starke, G., De Clercq, E., & Elger, B. S. (2021). Towards a pragmatist dealing with algorithmic bias in medical machine learning. *Medicine, Health Care and Philosophy*, *24*(3), 341–349. https://doi.org/10.1007/s11019-021-10008-5

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Strickland, E. (2019, October 24). Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients. Retrieved February 7, 2022, from IEEE Spectrum website: https://spectrum.ieee.org/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients

von Schomberg, R. (2013). A Vision of Responsible Research and Innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible Innovation* (pp. 51–74). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch3

Wen, D., Khan, S. M., Ji Xu, A., Ibrahim, H., Smith, L., Caballero, J., … Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: A systematic review. *The Lancet Digital Health*, *4*(1), e64–e74. https://doi.org/10.1016/S2589-7500(21)00252-1

Williams, J. V., & Razavian, N. (2019). *Towards Quantification of Bias in Machine Learning for Healthcare: A Case Study of Renal Failure Prediction*. 5.

Winnicki, E., McCulloch, C. E., Mitsnefes, M. M., Furth, S. L., Warady, B. A., & Ku, E. (2018). Use of the Kidney Failure Risk Equation to Determine the Risk of Progression to End-stage Renal Disease in Children With Chronic Kidney Disease. *JAMA Pediatrics*, *172*(2), 174–180. https://doi.org/10.1001/jamapediatrics.2017.4083

Yates, D. J., Gulati, G. J. J., & Weiss, J. W. (2013). Understanding the Impact of Policy, Regulation and Governance on Mobile Broadband Diffusion. *2013 46th Hawaii International Conference on System Sciences*, 2852–2861. Wailea, HI, USA: IEEE. https://doi.org/10.1109/HICSS.2013.583