Precision Medicine and the Machine Learning: Considerations Regarding Social Factors in Designing an Algorithm

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 of the Degree
Bachelor of Science, School of Engineering

Rayaan Faruqi
Spring, 2021

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

Signature ________________________________ Date: 04/27/2021
Rayaan Faruqi

Introduction

Machine learning (ML) and artificial intelligence (AI) can be racist. These algorithms can be sexist, homophobic, and otherwise biased against certain groups of people ("Finding Bias in Health Care AI Wins STAT Madness 'Editors' Pick'," 2020; "The Best Algorithms Still Struggle to Recognize Black Faces," n.d.). ML and what we can consider an evolving field of ML, AI, are fundamentally about defining mathematical relationships between data (Khan, 2020). If the data are skewed or biased, ML algorithms can learn dangerous prejudices that harm various demographics of people, which is especially problematic in the context of medicine.

Medicine is becoming increasingly patient-data focused, moving toward a paradigm of precision medicine, which is the development of procedures, techniques, and treatment plans that consider factors such as race, socioeconomic status, and various other social variables (Ho et al., 2020). Machine learning algorithms are well suited for precision medicine purposes because they are designed around developing highly complex relationships between enormous datasets. If the developers of these algorithms are not careful, these algorithms can fall prey to biases that harm certain patient groups.

The thesis applies Thomas Kuhn's paradigm shift and Sheila Jasanoff's co-production theories to determine if and how a paradigm shift from current medical techniques to precision medicine is occurring, and to examine how machine learning and medicine are co-producing one another. With the use of these frameworks, insights into how to encourage fairness and equity in future machine learning developments, and by extension in medicine, are developed. These insights serve to help answer the question: how should machine learning and medicine's mutual growth and enmeshment be directed to ensure equitable, fair, and effective treatment for all demographics?

Applying Frameworks: Paradigm Shift and Co-Production in Precision Medicine

Thomas Kuhn's paradigm shift and Sheila Jasanoff's co-production theories are used to explore and inform the question of how to ensure equitable development of precision medicine, a new paradigm within medicine, with machine learning. Kuhn defines the paradigm shift as a fundamental change from the status quo to a new mode of thinking entirely, analogous to how the view of the helium atom changes based upon considering it from a physical or chemical perspective (Kuhn, 1970). The paradigm shift is said to have happened when new techniques and modes of thinking render "normal science" obsolete (Kuhn, 1970). Jasanoff describes co-production as a system in which two variables affect one another simultaneously, and in doing so define and build the system together (Jasanoff, 2004). Both Kuhn's and Jasanoff's frameworks have received some criticism.

Criticisms of Kuhn's paradigm shift included suggestions that his conception of normal science is inherently false, that Kuhn fails to acknowledge and address the frequency with which revisions in "normal science" occur, and that his thesis of incommensurability between paradigms is vastly oversimplified (Dolby, 1971; Kordig, 1973; Toulmin, 1973). Jasanoff's work has, to a lesser extent, also received criticism, with regard to the vague parameters with which coproduction is defined (Filipe et al., 2017; Fukushima, 2013). With these critical considerations in mind, this report first applies the paradigm shift framework and second applies the co-production framework.

Kuhn's framework determines whether a paradigm shift is happening within the medical field. The primary paradigm shift to examine is that of the shift from typical modes of medical practice to those of precision medicine, a paradigm in which patient-specific genomic and social factors are used to refine and individualize treatment for maximum efficacy, especially using machine learning techniques. In other words, precision medicine is the refinement of treatments to consider more individual factors about the patient to construct the best treatment plan and prognoses possible (HealthITSecurity, 2018; *What Is the Difference between Precision Medicine and Personalized Medicine?*, n.d.). Criticisms of Kuhn's work do not apply in the interaction of medicine and machine learning. Though it is true that normal science is constantly subject to revision and that his theory is a simplified conception of how science evolves, it is also true that the current medical field is undergoing a drastic change from its status quo. Modern medicine will eventually be rendered obsolete and incompatible with new modes of thinking that are brought about by the continued interaction, or the coproduction relationship, between it and machine learning.

Applying the co-production framework will examine how machine learning and medicine are *currently* co-producing the paradigm shift to precision medicine, and how future developments should be guided to co-produce a medical paradigm that offers fairness and equality to as many populations as possible. This application also involves examining current machine learning methods, focusing on training and validation datasets that they use, and the use or nonuse of social and racial parameters in modeling. Jasanoff's critics are correct in that the parameterization of a paradigm shift is vague in some cases, but as her theory is applied here, the interaction between medicine and ML fit her description of how something can evolve through the cyclical interactions of factors within a given system.

Paradigm shift and co-production theories will be used in concert with various STS research methods to justify the need for careful guidance of ML in medicine and to generate recommendations for algorithm creators, as detailed in the **Methodology** section.

Literature Review

What are Machine Learning and Artificial Intelligence?

ML and AI have historically been relegated to the realm of science fiction, popularized by television programs, video games, and other media. Though ML and AI are currently unable to fully imitate a human, the technology is advancing at an incredible rate. Machine learning algorithms, broadly speaking, use three types of learning: supervised, unsupervised, and reinforcement (Khan, 2020). In spite of their differences, all machine learning algorithms may be described fundamentally as building a mathematical relationship between datasets. ML algorithms can produce powerful equations with vast numbers of factors in consideration, allowing whoever is using them unique insight into virtually any situation. Artificial intelligence can be considered an evolution of machine learning; AI is effectively the use of a combination of machine learning techniques to produce an immensely complex algorithm that can make human-like decisions, evaluations, and predictions. ML and AI have the potential to vastly improve complicated fields such as medicine. Before the technology is applied freely, however, engineers and innovators must navigate issues that machine learning can exacerbate or create.

Potential Problems with Machine Learning

Though ML and AI can revolutionize the field of medicine, it is important to consider the role of clinicians and engineers in guiding the evolution of ML/AI technology and ensuring that it is able to reach all people fairly and equitably. As these models become more complex, a potential consequence is that the algorithm becomes so difficult to decipher that we no longer know *how* results are being computed. Trying to understand the model would in essence be attempting to decipher a black box. This inability to validate and understand an algorithm has dangerous implications; what if a medical AI carries an unseen bias that causes it to perform worse for minority patients? What if an insurance algorithm disproportionately discriminates based on protected factors such as race, ethnicity, gender, or sexual orientation?

Machine learning algorithms have already demonstrated the ability to convey biases. For example, software designed to alert Nikon camera users of blinking subjects tends to "interpret Asians as always blinking" (Zou & Schiebinger, 2018). Commercial facial recognition systems misclassified gender at a rate of 0.8% for lighter-skinned men versus 35% for darker-skinned women (Buolamwini & Gebru, 2018). It is unlikely that the programs are specifically designed to be racist, but instead were built on homogenous datasets that failed to train the models to classify and calculate output equitably. An undecipherable algorithm that also carries inherent biases is a disastrous prospect for medical and surgical fields.

Recent papers have explored the potential use of ML and AI in driving surgery, analyzing patient health records, generating treatment plans, and predicting prognoses for patients (Adkins, 2017; Bihorac et al., 2019; Ho et al., 2020; Khan, 2020). Given the significance of what these developments could mean for the healthcare system, a series of ethical analyses have followed these papers, discussing questions one must employ in evaluating machine learning algorithms, considerations on structural racism, and more (Basu et al., 2020; Binns, 2018; Robinson et al., 2020; Vollmer et al., 2018).

These papers all point to a paradigm shift in the way that medicine is researched, conducted, and thought about; medicine is becoming more personalized. As medicine becomes more specifically tailored to individual differences in genome, race, socioeconomic status, sex, a newer and enhanced set of techniques will appear, going by the name precision medicine. For example, Lee et. al in a 2018 paper used genomic data as a basis for the identification of reliable gene expression markers for drug sensitivity in the treatment of acute myeloid leukemia (Lee et al., 2018). This refinement of acute myeloid leukemia treatments is a part of a trend that will lead to the paradigm shift of medicine to precision medicine. This paradigm shift is coproduced by the interaction of medicine, machine learning, and stakeholders – clinicians, engineers, scientists, patients, hospitals, and regulatory agencies.

Machine learning is well-suited to enormously complex tasks like those that involve the consideration of many hundreds, or even thousands of variables. It is conceivable soon an ML algorithm will be able to diagnose and form treatment plans for patients. It is also conceivable that they will become far better at it than a human being ever will be, given their ability to mathematically describe immensely complex relationships. In being able to consider factors that lend to precision medicine, like race, socioeconomic status, sex, genomic data, as well as data from hundreds or thousands of patient cases, an ML algorithm may make a better doctor than any human ever could. Eventually, it may be a full-blown artificial intelligence that is creating treatment plans and even performing surgery. While this could usher in a new era of medical success, these algorithms may become so complex that we are no longer able to understand them.

In exploring how the paradigm of medicine is shifting to that of precision medicine, it is useful to explore how we may be able to guide coproduction at a micro and a macro level. Khan et al. for instance stresses the importance of having a human being work in parallel with a machine learning model to drive personalized medicine approaches for spinal care (Khan, 2020). Jasanoff almost describes co-production this way: as a system in which two variables affect one another simultaneously, and in doing so define the system together (Jasanoff, 2004).

Methodology

The thesis addresses how machine learning and medicine's mutual growth and enmeshment should be directed to ensure equitable, fair, and effective treatment for all demographics.

As established in the literature review, a paradigm shift is being co-produced by the intersectional advancement of medicine and machine learning. Documentary analysis will explore literature that involves the inclusion/exclusion of social factors in algorithmic parameters and potential solutions that other scientists are arriving at in an effort to address ethical issues arising with ML and medicine. Additionally, I will perform a historical analysis to examine which social factors/algorithmic parameters carry the most weight in ensuring equitable model performance.

Results and Discussion

When speaking of machine learning and datasets, we must consider both the quantity and quality of a dataset. Machine learning processes are largely statistical in nature, in that they use raw data as the basis for building relationships. ML largely may be thought of as an evolution and automation of data analysis. For this reason, the first recommendation is to include as much data as possible in order to arrive at mathematical relationships between factors with the highest possible levels of confidence. As a rule of thumb, "the more data, the better."

The quantity of data is not all that matters. The data input into a machine learning model must be diverse and capture a wide variety of features from many different groups of people. A lack of contextual specificity, referring to the availability of data for different demographics, skews the datasets and therefore the results associated with a particular ML algorithm (Panch et al., n.d.). We have already seen the effects of this with a photography algorithm disproportionately interpreting Asians as blinking, and a commercial face recognition system misclassifying sex at a rate of 0.8% for lighter-skinned males and 35% for darker-skinned females (Buolamwini & Gebru, 2018; Zou & Schiebinger, 2018). To this end, the second recommendation is to maximize the diversity of training datasets and include data from and for as many protected groups as possible, striving for a standard of "algorithmic inclusivity" (Faruqi & Singh, 2021).

However, this standard of algorithmic inclusivity goes beyond merely collecting data from those protected groups. It also refers to the explicit inclusion of social factors such as race, socioeconomic status, and sexual orientation. Fairness via the principle of unawareness, or the explicit exclusion of social factors, is recommended against, as complex models are capable of inferring those factors (Rajkomar et al., 2018). For example, it was discovered that an algorithm for the prediction of cardiovascular risk factors was able to infer the self-reported sex of a patient based on retinal fundus photographs (Poplin et al., 2018). Additionally, in the cases that certain features do not matter to the performance of an algorithm, a machine learning model will deemphasize their importance in the parameterization schema, a process that only becomes more effective when the principles of dataset quality and quantity are adhered to.

In 2018, Kleinberg et al. tested the inclusion of race in an algorithm to predict student success in college. They found that not only did the inclusion of race improve the predicted GPAs of admitted students, but that it could also improve outcomes such as the proportion of black students admitted (Kleinberg et al., 2018). Another paper found that the *modus operandi* of anti-classification, or striving for fairness through algorithmic blindness to protected traits, suffers from deep statistical limitations and is able to harm both minority and majority communities (Corbett-Davies & Goel, 2018). As an example, completely gender-neutral risk scores in the criminal justice system may encourage unnecessarily harsh sentencing for women.

It can also be demonstrated more generally that the exclusion of any information, including protected attributes, can lead to discriminatory decisions (Corbett-Davies & Goel, 2018). In a synthetic data experiment, a case study, and examples of public datasets, Lipton et al. also discovered that they were able to achieve *impact parity* (equitable outcomes) via *treatment disparity* (treating members of protected subgroups differently) (Lipton et al., 2019). For historical, albeit non-ML, counterexamples, we may look at how equating blindness with fairness is deeply problematic. For example, sex-blind orchestra auditions failed to combat implicit bias against women candidates in the 1950s, and zip codes may be used as a proxy for race in the United States, as seen in the finance and banking industry's redlining practices (*How Blind Auditions Help Orchestras to Eliminate Gender Bias*, 2013; Abuhamad, 2019).

Unfortunately, literature that explicitly tests the idea of algorithmic fairness via blindness to protected traits is scarce. At the time of this writing and to my knowledge, Kleinberg et al., Corbett-Davies & Goel, and Lipton et al. have been the only researchers to do so. Though it can reasonably be recommended, based on the above discussion, that social factors like race and sex should be included in algorithms, more research is required for the broader scientific acceptance that protected factors *should* be included in an ML algorithm.

A final recommendation is to encourage a culture and tendency among companies, engineers, and scientists for transparency in the algorithms they write. In other words, we must shift the proverbial needle to a white-box mentality. This would allow for greater trust in the algorithms scientists and engineers create, and would allow for regulators and other stakeholders to audit a given algorithm (Halter et al., 2009; Hyndman & McConville, 2016; Islam et al., 2002; Weller, 2019). As recommended by myself and Singh in a previous corporate publication, this may be done via comprehensible explanations of general algorithmic function, visualizations of input/outputs model architecture, and measures of which parameters are given most importance when arriving at a particular output or classification (Faruqi & Singh, 2021).

Implementing these strategies is more difficult in practice, as there are a host of reasons a company may choose not to divulge algorithmic specifics (e.g. intellectual property, data protection and privacy). Such transparency practices, however, can be made more palatable via data anonymization techniques, coarsening precision values by rounding prediction probabilities to fewer digits, increasing variance and uncertainty by modifying prediction vector entropy, and employing regularization techniques to avoid overfitting (Faruqi & Singh, 2021; Jia & Gong, 2020; Mozaffari-Kermani et al., 2015; Rudin, 2019; Yeom et al., 2018).

Conclusion

Machine learning and artificial intelligence have the power to change medicine for the better. With these algorithms' incredible ability to transform enormous amounts of data into meaningful and accurate relationships, precision medicine may become the norm, with medical technology and treatment efficacy reaching heights never seen before. In spite of this hopefulness, it is important the medical machine learning field remain vigilant and move forward deliberately but prudently. Machine learning and artificial intelligence are in their current forms far removed from any humanity, but their abilities and aptitude for biases reflect ours all too well. Engineers and scientists should carefully vet the quality and quantity of our datasets, as data is the foundation of all machine learning algorithms. They should strive for standards of algorithmic inclusivity and give algorithms the data they need so that they may serve all people to their best ability. The writers of machine learning algorithms should lean into a culture of transparency and a white-box mentality, so that their work and its impact are more easily corrected and audited when it falls prey to the insidious effects of bias. More research must also be conducted in the sphere of algorithmic fairness with respect to the specific inclusion or exclusion of protected attributes, as this will help encourage both awareness and understanding of the idea that we might reach impact parity via the counterintuitive application of treatment disparity. We must strive to achieve the aforementioned values because if we do not, the future of machine learning and medicine will not be one of equality, equity, order, and of service to all peoples. It may instead become a future of unequal and uncontrollable algorithms, working to service the few and the privileged, where the historically marginalized and oppressed grow more distant from equity, equality, and justice.

References

Abuhamad, G. M. (Grace M. (2019). *The fallacy of equating "blindness" with fairness: Ensuring trust in machine learning applications to consumer credit* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/122094

Adkins, D. E. (2017). Machine Learning and Electronic Health Records: A Paradigm Shift. *American Journal of Psychiatry*, *174*(2), 93–94. https://doi.org/10.1176/appi.ajp.2016.16101169

Basu, T., Engel-Wolf, S., & Menzer, O. (2020). The ethics of machine learning in medical sciences: Where do we stand today? *Indian Journal of Dermatology*, *65*(5), 358–364. https://doi.org/10.4103/ijd.IJD_419_20

Bihorac, A., Ozrazgat-Baslanti, T., Ebadi, A., Motaei, A., Madkour, M., Pardalos, P. M., Lipori, G., Hogan, W. R., Efron, P. A., Moore, F., Moldawer, L. L., Wang, D. Z., Hobson, C. E., Rashidi, P., Li, X., & Momcilovic, P. (2019). MySurgeryRisk: Development and Validation of a Machine-Learning Risk Algorithm for Major Complications and Death after Surgery. *Annals of Surgery*, *269*(4), 652–662. https://doi.org/10.1097/SLA.0000000000002706

Binns, R. (2018). *Fairness in Machine Learning: Lessons from Political Philosophy*. 11.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Corbett-Davies, S., & Goel, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *ArXiv:1808.00023 [Cs]*. http://arxiv.org/abs/1808.00023

Dolby, R. G. A. (1971). Philosophy of Science - Criticism and the Growth of Knowledge. Proceedings of the International Colloquium in the Philosophy of Science, London 1965, Volume 4. Ed. by Imre Lakatos and Alan Musgrave. London: Cambridge University Press. Pp. viii + 282. 1970. £3.50. *The British Journal for the History of Science*, *5*(4), 400–400. https://doi.org/10.1017/S0007087400011626

Faruqi, R., & Singh, A. (2021). *Best Practices for Addressing Risks Associated with a Lack of Diversity in Machine Learning*. 12.

Filipe, A., Renedo, A., & Marston, C. (2017). The co-production of what? Knowledge, values, and social relations in health care. *PLOS Biology*, *15*(5), e2001403. https://doi.org/10.1371/journal.pbio.2001403

Finding bias in health care AI wins STAT Madness "Editors" Pick'. (2020, April 6). *STAT*. https://www.statnews.com/2020/04/06/stat-madness-editors-pick-racial-bias-in-health-care-ai/

Fukushima, M. (2013). Between the Laboratory and the Policy Process: Research, Scientific Community, and Administration in Japan's Chemical Biology. *East Asian Science, Technology and Society*, *7*(1), 7–33. https://doi.org/10.1215/18752160-2074955

Halter, M. V., de Arruda, M. C. C., & Halter, R. B. (2009). Transparency to Reduce Corruption? *Journal of Business Ethics*, *84*(3), 373. https://doi.org/10.1007/s10551-009-0198-6

HealthITSecurity. (2018, January 11). *What Are Precision Medicine and Personalized Medicine?* HealthITAnalytics. https://healthitanalytics.com/features/what-are-precision-medicine-and-personalized-medicine

Ho, D., Quake, S. R., McCabe, E. R. B., Chng, W. J., Chow, E. K., Ding, X., Gelb, B. D., Ginsburg, G. S., Hassenstab, J., Ho, C.-M., Mobley, W. C., Nolan, G. P., Rosen, S. T., Tan, P., Yen, Y., & Zarrinpar, A. (2020). Enabling Technologies for Personalized and Precision Medicine. *Trends in Biotechnology*, *38*(5), 497–518. https://doi.org/10.1016/j.tibtech.2019.12.021

How blind auditions help orchestras to eliminate gender bias. (2013, October 14). The Guardian. http://www.theguardian.com/women-in-leadership/2013/oct/14/blind-auditions-orchestras-gender-bias

Hyndman, N., & McConville, D. (2016). Transparency in Reporting on Charities' Efficiency: A Framework for Analysis. *Nonprofit and Voluntary Sector Quarterly*, *45*(4), 844–865. https://doi.org/10.1177/0899764015603205

Islam, R., Djankov, S., & McLeish, C. (2002). *The Right to Tell*. The World Bank. https://doi.org/10.1596/0-8213-5203-2

Jasanoff, S. (2004). *States of Knowledge: The Co-Production of Science and the Social Order*. Routledge.

Jia, J., & Gong, N. Z. (2020). AttriGuard: A Practical Defense Against Attribute Inference Attacks via Adversarial Machine Learning. *ArXiv:1805.04810 [Cs, Stat]*. http://arxiv.org/abs/1805.04810

Khan, O. (2020). Use of Machine Learning and Artificial Intelligence to Drive Personalized Medicine Approaches for Spine Care. *WORLD NEUROSURGERY*, 7.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic Fairness. *AEA Papers and Proceedings*, *108*, 22–27. https://www.jstor.org/stable/26452698

Kordig, C. R. (1973). Observational Invariance. *Philosophy of Science*, *40*(4), 558–569. https://www.jstor.org/stable/186288

Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl). University of Chicago Press.

Lee, S.-I., Celik, S., Logsdon, B. A., Lundberg, S. M., Martins, T. J., Oehler, V. G., Estey, E. H., Miller, C. P., Chien, S., Dai, J., Saxena, A., Blau, C. A., & Becker, P. S. (2018). A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nature Communications*, *9*(1), 42. https://doi.org/10.1038/s41467-017-02465-5

Lipton, Z. C., Chouldechova, A., & McAuley, J. (2019). Does mitigating ML's impact disparity require treatment disparity? *ArXiv:1711.07076 [Cs, Stat]*. http://arxiv.org/abs/1711.07076

Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., & Jha, N. K. (2015). Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare. *IEEE Journal of Biomedical and Health Informatics*, *19*(6), 1893–1905. https://doi.org/10.1109/JBHI.2014.2344095

Panch, T., Mattie, H., & Atun, R. (n.d.). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health*, *9*(2). https://doi.org/10.7189/jogh.09.020318

Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, *2*(3), 158–164. https://doi.org/10.1038/s41551-018-0195-0

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine*, *169*(12), 866–872. https://doi.org/10.7326/M18-1990

Robinson, W. R., Renson, A., & Naimi, A. I. (2020). Teaching yourself about structural racism will improve your machine learning. *Biostatistics*, *21*(2), 339–344. https://doi.org/10.1093/biostatistics/kxz040

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

The Best Algorithms Still Struggle to Recognize Black Faces. (n.d.). *Wired*. Retrieved October 27, 2020, from https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/

Toulmin, S. (1973). Human Understanding, Vol. I: The Collective Use and Evolution of Concepts. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie*, *4*(2), 398–402.

Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K. S. L., Myles, P., Granger, D., Birse, M., Branson, R., Moons, K. G., Collins, G. S., Ioannidis, J. P. A., Holmes, C., & Hemingway, H. (2018). Machine learning and AI research for Patient Benefit: 20 Critical Questions on Transparency, Replicability, Ethics and Effectiveness. *ArXiv:1812.10404 [Cs, Stat]*. http://arxiv.org/abs/1812.10404

Weller, A. (2019). Transparency: Motivations and Challenges. *ArXiv:1708.01870 [Cs]*. http://arxiv.org/abs/1708.01870

What is the difference between precision medicine and personalized medicine? What about pharmacogenomics?: MedlinePlus Genetics. (n.d.). Retrieved October 14, 2020, from https://medlineplus.gov/genetics/understanding/precisionmedicine/precisionvspersonalized/

Yeom, S., Giacomelli, I., Fredrikson, M., & Jha, S. (2018). Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting. *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, 268–282. https://doi.org/10.1109/CSF.2018.00027

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, *559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8