IMPACT OF UNETHICAL BIASES ON RECOMMENDER SYSTEMS

A Research Paper submitted to the Department of Engineering and Society
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By

Michael Bassilios

March 25, 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.

ADVISOR
Catherine D. Baritaud, Department of Engineering and Society

Currently, the college golf recruiting process is very tedious and fragmented for both college coaches and high school golfers, many of whom are international students. My technical project consists of determining factors that make a player and school a good fit, analyzing datasets to build predictive models, and ultimately including these factors and models in a robust, online recommender system that can be used by players and coaches. This system that has two main tasks: i) to give college coaches a list of potential junior golfers who may fit on their team and ii) to provide recommendations to the junior golfers about which school may be a good fit for them.

This tightly coupled STS research focuses on the biases that exist in recommender systems. Companies use AI hiring tools, such as recommender systems, in an effort to make their hiring or recruiting processes fairer. Since all humans come with some amount of bias, companies have hoped that the use of these tools would solve the issue of bias in recruiting and hiring, resulting in a more unbiased process. However, these systems do not necessarily make the process fairer, and in some cases, may even make it worse. What complicates matters further is that, since recommender systems are developed by humans, and all humans have bias, then the recommender system will naturally contain human bias in it. The concern of bias in these systems is extremely urgent because, over the past few years, AI hiring tools have become more widespread and companies have been increasingly reliant on them. Additionally, many people do not realize that these tools contain bias because they may not be aware of their own biases to begin with. There are various types of biases that can affect recommender systems, which result in detrimental impacts on society as whole, particularly on both the recruit and recruiter. However, not all hope is lost for recommender systems and other AI hiring tools. Some solutions exist that can potentially reduce bias significantly in recommender systems, ultimately making the recruiting process fairer. The Actor-Network Theory framework will be used to examine the social groups, problems, and solutions involving recommender systems.

TYPES OF UNETHICAL BIASES IN RECOMMENDER SYSTEMS

Over time, a variety of biases have been introduced to recommender systems, and they continue to plague the systems and, therefore, the recruiting process. These biases can come in different forms, but the commonality between them all is that they diminish the effectiveness of the recommender systems. Some of the more prominent types of biases in these systems include: racial, gender, age, sexual, popularity, and disability. While society generally understands the meaning of each of these biases, they are often difficult to recognize when they are contained in the system itself. For instance, popularity bias is defined by Abdollahpour, Burke, and Mobasher (2017) as "collaborative filtering recommenders emphasize popular items…much more than other [items]" (p. 42). In regard to disability bias, there have been studies where "researchers found that…a machine-learned sentiment analysis model rates texts which mention disability as more negative" (Whittaker et al., 2019, p. 8). Age bias, or ageism, can enter a system when a company tends to focus on hiring candidates that are around a particular age or have a certain amount of experience, such as candidates that are younger or are recent graduates. Clearly, all of these biases in the system can affect certain groups in society in different ways, so it is extremely important for society as a whole to be aware of two points when it comes to bias in recommender systems. The first point is that people must be aware of some biases that they may have. The second is that they must realize that their biases are consistent with those found in the systems. The former point presents a challenge, however, because many of people's biases are unintentional, making them difficult to perceive. One example that displays unconscious bias comes from the Race Implicit Association Test (IAT), which shows one's racial bias. According to Marcelin, Siraj, Victor, Kotadia, and Maldonado (2019), this test was taken over four million times from 2002 to 2017, and 75% of participants showed automatic white preference, which associates white people with good and black people with bad (p. S64). The latter point also presents a challenge in a slightly different way. Not only do all people have biases, many of which are unconscious, but "people have a lot of different and unusual biases, which is fascinating but also quite terrifying" (Fleming, 2019, para. 14). Therefore, since each person comes with varying types and levels of biases, the biases in the systems will also vary, making them tougher to recognize.

IMPACT OF UNETHICAL BIASES ON RECOMMENDER SYSTEMS

Given all of the biases that can contaminate recommender systems, it is more important to consider their impact on the recruiting process as a whole. Adomavicius, Bockstedt, Curley, and Jingjing (2019) state that "biases can contaminate the recommender system's inputs, weakening the system's ability to provide high-quality recommendations" (p. 1322). This is a significant problem because these biases can ultimately be a determining factor for which candidate gets selected by the recruiter. If one candidate is better than another, but that candidate is negatively impacted through bias in the system, he or she may not be recommended by the system, which can result in a different candidate ultimately being selected for the position. Unfortunately, there are multiple examples and significant evidence of bias in recommender systems adversely impacting the recruiting process. Hsu (2020) provides a clear example of gender bias when Amazon recently stopped using an AI hiring tool because the system, "had learned to prefer male job candidates while penalizing female applicants" (p. 9). Hsu goes on to explain that this happened because the AI training originally consisted of mostly male candidate applications. This improper AI training is detrimental because the best candidate in this case may have been a woman, but she might not have gotten the job because the system is biased in favor of men. More gender bias was seen in a different way on LinkedIn's search engine a few years ago. Day (2016) explains that, "searches of popular female first names…bring up LinkedIn's suggestion to change 'Andrea Jones' to 'Andrew Jones,' 'Danielle' to 'Daniel,' 'Michaela' to 'Michael,' and 'Alexa' to 'Alex'" (para. 4). Day goes on to explain that the opposite did not happen, meaning that if one were to search 'Daniel', there would be no suggestion to change to 'Danielle'. These two examples not only show clear evidence of bias in recommender systems, but also display the detrimental effects they can have on recruits and recruiters. These effects are especially urgent considering that this technology is continuing to be used more and more by many companies. Amazon and LinkedIn are two very well-known companies that have had significant issues with controlling bias in their recommender systems, and they are not the only companies to have this problem. If recommender systems are not making recruiting processes any fairer, or even make them worse, then is there really a point to using them at all?

RECOMMENDER SYSTEMS THROUGH THE LENS OF ACTOR-NETWORK THEORY

A good framework for looking at recommender systems and ensuring their impartiality and success in the long run is Actor-Network Theory (ANT), which breaks down a technology into its social groups, problems faced by those groups, and possible solutions (Law and Callon, 1988). A benefit of using ANT is that it focuses on the entire network, which can be composed of both human and non-human actors. This framework will help illustrate some of the points and ideas that have been discussed so far. First, multiple social groups can influence or be influenced by the recommender system, and these groups are shown in Figure 1. For this cause-and-effect

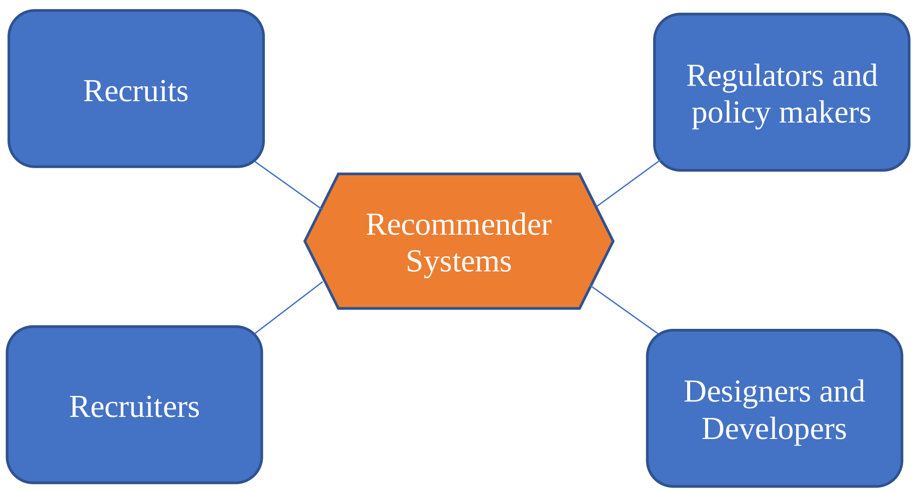

Figure 1: Relevant Social Groups for Recommender Systems. Four main actors who in some way impact or shape the recommender system (Adapted by Bassilios (2020) from Carlson 2009).

paper, designers and developers will be the main focus group, as that most closely ties in to the technical project, which involves developing a recommender system. Figure 1 can then be broken down into the different problems that designers and developers face, and those can be seen in Figure 2 on page six. These problems can range from issues with the system itself, to potential conflicts with other actors in the network, to negative effects of using the system. The issue of bias in the systems will be the main focus for this article. This is a complicated issue to solve because of the way in which bias enters the system originally. Buranyi (2017) puts it bluntly: “Computers don’t become biased on their own. They need to learn that from us” (para.

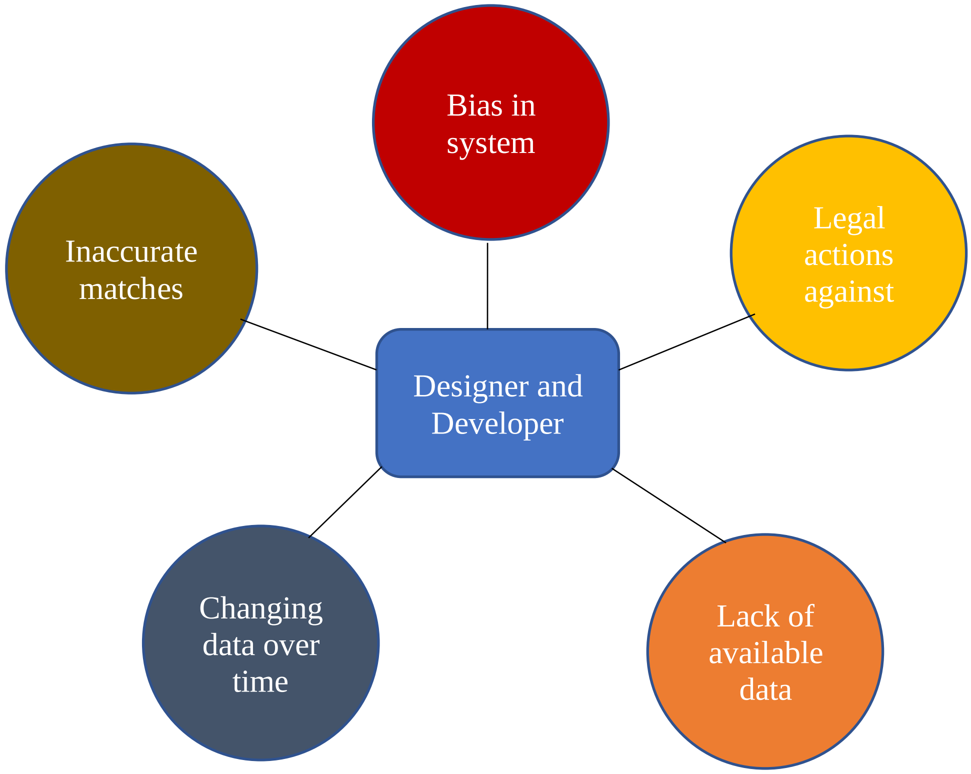

Figure 2: Potential Problems for a Social Group. Five possible issues that designers and develops, a main social group, may face when forming recommender systems (Adapted by Bassilios (2020) from Carlson 2009).

10). This is a strong statement from Buranyi, and it shows the impact that developers have on the system, as well as the importance of the input and data that they feed the system. She explains it further by saying, "as the algorithms learn and adapt from their original coding, they become more opaque and less predictable" (Buranyi, 2017, para. 11). Much of the responsibility of the 'original coding' that Buranyi mentions lies on the shoulders of the developers and designers, and can determine the amount of bias the system will have. Not all hope is lost, however, as Figure 3 on page seven provides some solutions for the problem of bias in the system. Using ANT to look at recommender systems is a strong method of analyzing the relationship between different actors in the network. The various actors will naturally have different or even conflicting ideas about how a recommender system should be developed. These varying opinions lead to the concept of framing, which is defined by Jolivet and Heiskanen (2010) as, "the process

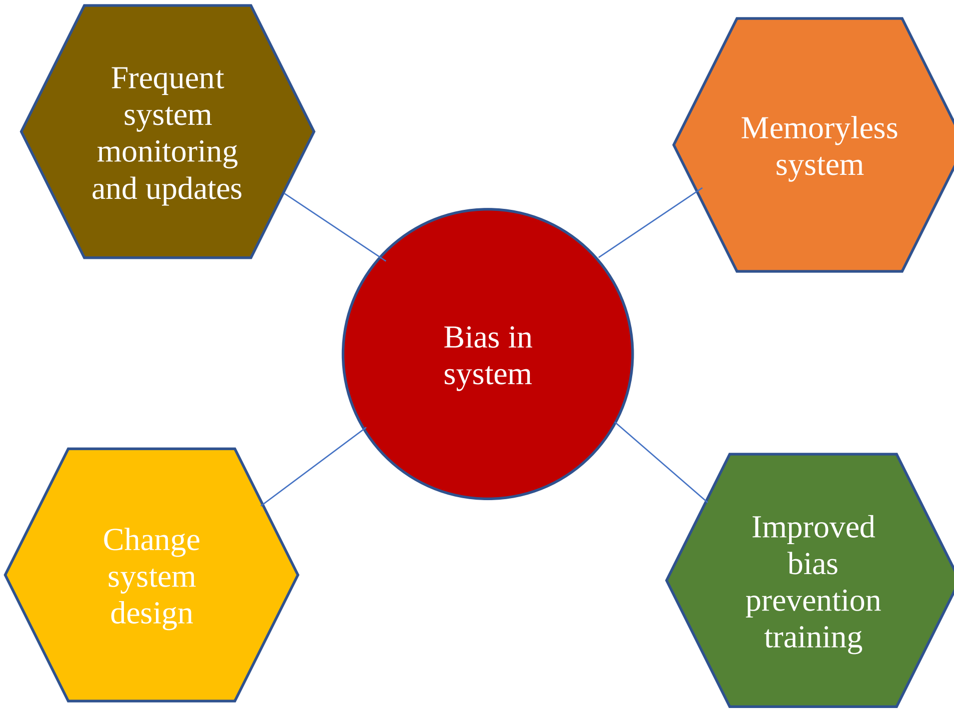

Figure 3: Potential Solutions for a Recommender System Problem. Four possible, general solutions to bias in the system, a major issue that its designers and developers face (Adapted by Bassilios (2020) from Carlson 2009).

through which a common world is established between different actors that allows them to achieve a collective scenario of a desired outcome" (p. 6748). Thus, framing is an integral part of the development and use of recommender systems when viewed through the lens of Actor-Network Theory because it requires the actors to come together, negotiate, and compromise in order to find a common ground.

SOLUTIONS TO REDUCE BIAS IN RECOMMENDER SYSTEMS

SOLUTION ONE: CHANGE SYSTEM DESIGN

The first solution is to change the design of the recommender system. Of the four solutions in Figure 3, this one would be the least effective for recommender systems used in the recruiting process, which is the focus of both this article and the technical project. However, this solution would be more valid for recommender systems used by one user, such as a search engine that recommends products.

SOLUTION TWO: MORE MEMORYLESS RECOMMENDER SYSTEM

The second solution would be to make the recommender system as memoryless as possible. While it is impossible to make the system fully memoryless because it needs input and training data, the goal here is to make each recommendation as independent as possible from any previous recommendations. In this way, one recommendation is not impacted by any prior one. The trouble with using non-memoryless recommender systems is that a biased recommendation would negatively affect all subsequent ones. Furthermore, “it can sometimes not make a prediction for certain active users…if the active user has no items in common…” (Carlton College, n.d.). Therefore, by making the system as memoryless as possible, the bias in future recommendations can be reduced, and groups would not be automatically overlooked by the system, even if they are less common.

SOLUTION THREE: MORE SYSTEM UPDATES AND MONITORING

The third solution is to have more frequent system monitoring and updates. One of the biggest issues that plague recommender systems is that after they are developed and released for use, they are not monitored nearly as much as they should be. Companies and recruits tend to blindly trust the recommendations that the system makes without ensuring that the system is actually effective and impartial. As Adomavicius et al. and Buranyi hinted at earlier, the biases in the system can increase and become worse and less predictable over time. Generally, any technology should be monitored and updated closely, particularly those that involve AI. For instance, smartphones and operating systems have frequent updates to fix bugs and various issues that may arise. It should be no different for recommender systems, since they are significantly affected by biases. McLaren (2019) is very straightforward when she says, "The most important step any company can take when adopting a new AI tool is to closely monitor the results it produces" (para. 3). By taking this step, companies can quickly realize if any changes or updates are needed to improve the system's performance. Unfortunately, companies and developers tend to wait until a clear problem emerges before making changes, and this was clearly the case in the Amazon and LinkedIn examples earlier. Biases may still make their way into the system, but it is extremely crucial that they are caught early so that they do not play a harmful role in the recruiting process over time. The best way to ensure that is through constant monitoring and frequent updates.

SOLUTION FOUR: IMPROVE BIAS PREVENTION TRAINING

To this point, none of the solutions have provided a way to completely rid the recommender systems of bias. Rather, they have focused on reducing bias in the systems and, more importantly, reducing the impact of the biases on the recruiting process in order to make the process fairer. The fourth and final solution is the most effective and logical one, but also the toughest one to accomplish. It is to provide more rigorous and personalized training in bias prevention, particularly focused on developers and recruiters. The goal of this solution is to reduce human bias. This is the most effective solution because human bias is the very source of the bias that exists in recommender systems. Thus, reducing human bias would automatically decrease the bias in the systems. While many companies currently have some sort of bias prevention training, these trainings simply do not do enough or go deep enough to solve the comlex issue of human bias. “Two-thirds of human resources specialists report that diversity training does not have positive effects” and “there is ample evidence that training alone does not change attitudes or behavior, or not by much and not for long” (Dobbin & Kalev, 2018, p. 49). Having one general, identical bias prevention training for all is not ideal for a couple of reasons. First, it is simply not rigorous or frequent enough, meaning that they are easily forgotten. In fact, “we forget 60% of what we’ve learned in 24 hours if it’s not reinforced. And awareness doesn’t lead to action” (Young, 2018, para. 9). Second, as cited by Fleming earlier, one person’s bias, whether conscious or unconscious, will likely be different from another’s. Since it has been established that people are unaware of their unconscious bias, it is much harder for them to connect to the training if they do not realize that bias is an issue for them in the first place. Simply put, the training can just become something for employees to check off of their to-do list. Some steps must be taken to address these shortcomings of current bias prevention training. First, it is imperative to show people their bias and make them aware of it. This can be done through some bias tests that currently exist, such as the Race IAT. Being exposed to their own biases may be uncomfortable for many people at first, and the results need not be shared, but it is a pivotal first step toward overcoming them. Then, when they see which biases they have, there must be specific prevention training geared toward those biases. Finally, there must be consistent, frequent training and testing so that people may see their own improvement over time. Bendrick and Nunes (2012) suggest that “posttraining testing could be used to assess whether training is effective” (p. 255). Frequent training is especially essential for employees like developers of recommender systems and recruiters since their bias may play a greater role in recruiting. By taking these steps, the bias prevention training is both more rigorous and more personalized. Admittedly, it may seem as though a substantial amount of effort is required from multiple groups, and that is true. This solution depends mainly on the willingness of employees, recruiters, and developers to go through a more rigorous training program. Still, it is needed because bias is a serious problem that gets worse if left unattended. Of the four given solutions, this one is the best for the long run and the future of recommender systems. It is vital to remember that the issue of bias in recommender systems is not the systems' fault, but that it stems from humans, and this solution takes that into account. The truth is that, since everyone has some level of bias, bias is not only seen in recommender systems, but in many aspects of the workplace and society in general. This solution not only improves the performance and impartiality of recommender systems, making the recruiting process fairer, but it will also have a positive impact on the workplace and society.

REVIEW AND RECAP

While the majority of this paper has looked at the role of bias in recommender systems and the harmful impact they have on the recruiting process, hope still remains for these systems, as seen in the suggested solutions. Recommender systems still have the potential to be a very useful technology that can have a positive impact on both the recruit and the recruiter in the long run. In order to reach the full potential of these systems, the bias that is in them simply must decrease significantly. That will only happen through a consistent and conscious effort from all social groups involved to combat and overcome natural human bias. In fact, most of the provided solutions, particularly the last one, depend on these consistent efforts from various social groups; that is the only way that those solutions will be possible. Ultimately, the main takeaway is that the primary goal of recommender systems should not be to save time or make the recruiting process simpler. While those are valuable goals, they should be secondary. The primary goal must be to make the recruiting process fairer by providing an accurate, unbiased match between the recruit and recruiter. Doing so will be beneficial to the recruit since he or she gets to be a part of a fairer recruiting process, and it will be beneficial for the company, since it gets to truly hire the best possible candidate.

There are plenty of options for future work relating to this topic of bias in recommender systems. One of the most important next steps will be to analyze the implementation time and monetary cost necessary for each of the given solutions. Specifically, it is critical to look at the first three solutions from an economic and policy standpoint, while the fourth solution needs to examined from a psychological viewpoint. Additionally, data collection will certainly be an essential next step for all of the provided solutions, especially in checking their effectiveness over time. For example, data from the third solution could be analyzing the frequency of updates and monitoring on the diversity of recommendations. Data that can be collected in the fourth solution is the average change in bias test scores for employees over time while maintaining employee confidentiality, which can show the effectiveness of the more rigorous and personalized training. Data collection will help ensure that the bias level in the systems are actually decreasing over time, and can hopefully show and quantify the positive impact of the solutions on the recruiting process.

REFERENCES

Abdollahpouri, H., Burke, R., & Mobasher, B. (2017). Controlling popularity bias in learning-to-rank recommendation. In *Research Gate,* 42-46. doi: http://dx.doi.org/10.1145/3109859.3109912

Adomavicius, G., Bockstedt, J., Curley, S., & Jingjing, Z. (2019, December 1). Reducing recommender system biases: an investigation of rating display designs. *MIS quarterly* 43(4), 1321-1339. doi: 10.25300/MISQ/2019/13949

Bassilios, M. (2020, October 15). *Relevant Social Groups for Recommender Systems*. [Figure 1]. *STS Research Paper: Impact of Unethical Biases on Recommender Systems* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bassilios, M. (2020, October 15). *Potential Problems for a Social Group*. [Figure 2]. *STS Research Paper: Impact of Unethical Biases on Recommender Systems* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bassilios, M. (2020, October 15). *Potential Solutions for a Recommender System Problem*. [Figure 3]. *STS Research Paper: Impact of Unethical Biases on Recommender Systems* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Bendick Jr., M. & Nunes, A. (2012). Developing the research basis for controlling bias in hiring. *Journal of Social Issues*, 68(2), 238-262. doi: 10.1111/j.1540-4560.2012.01747.x

Buranyi, S. (2017, August 8). Rise of the racist robots – how AI is learning all our worst impulses. *The Guardian.* Retrieved from https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses

Carlton College. (n.d.). Recommender systems. In *A computer science comprehensive exercise.* Retrieved from https://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/memorybased.html

Day, M. (2016, August 31). How LinkedIn's search engine may reflect a gender bias. *The Seattle Times.* Retrieved from https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias/

Dobbin, F. & Kalev, A. (2018). Why doesn't diversity training work?. *Anthropology Now,* 10(2), 48-55. doi: 10.1080/19428200.2018.1493182

Fleming, N. (2019, January 12). Everyone is biased, including you: the play by neuroscientists. *The Gaurdian.* Retrieved from https://www.theguardian.com/science/2019/jan/12/psychology-of-group-reasoning-versus-individual

Hsu, J. (2020, September). Can AI hiring systems be made antiracist? *IEEE Spectrum,* 57(9), 9-11. doi: 10.1109/mspec.2020.9173891

Jolivet, E. & Heiskanen, E. (2010). Blowing against the wind—an exploratory application of actor network theory to the analysis of local controversies and participation processes in wind energy. *Energy Policy* 38, 6746-6754. doi: 10.1016/j.enpol.2010.06.044

Law, J. & Callon, M. (1988). Engineering and sociology in a military aircraft project: a network analysis of technological change. *Social Problems* 35(3), 284-297. doi: 10.2307/800623

Marcelin, J., Siraj, D., Victor, R., Kotadia, S., Maldonado, Y. (2019, August 20). The impact of unconscious bias in healthcare: how to recognize and mitigate it. In *The Journal of Infectious Diseases* 220(2), S62-S73. Retrieved from https://academic.oup.com/jid/article/220/Supplement_2/S62/5552356

McLaren, S. (2019, April 9). Keep bias out of your AI hiring tools with these 3 steps. *LinkedIn.* Retrieved from https://business.linkedin.com/talent-solutions/blog/artificial-intelligence/2019/keep-bias-out-of-ai-hiring-tools-with-3-steps

Whittaker, M., Alper, M., Bennett, C., Hendren, S., Kaziunas, L., Mills, M., … West, M. (2019). Disability, bias, and AI. *AI Now*, 1-32. Retrieved from https://ainowinstitute.org/disabilitybiasai-2019.pdf

Young, F. (2018). If unconscious bias training doesn't work, what's the alternative?. *Hive Learning.* Retrieved from https://www.hivelearning.com/site/if-unconscious-bias-training-doesnt-work-whats-the-alternative/