MEMORANDUM

May 3, 2021

TO: The ASE 333T Problem Statement Report Supervisor

PREPARED BY: Justin Campbell, Oliver Coyle, Katharine Fisher, & Nischal Shrestha

Racial Biases in Facial Recognition Technology

ABSTRACT

The purpose of this report is to describe the causes, consequences, and possible solutions to racial biases in facial recognition systems used by law enforcement. Facial recognition systems analyze an image of a face and infer information about the individual or match the face to an image in a database. These systems demonstrate increased inaccuracy when identifying racial minorities. When law enforcement uses this technology, false matches can lead to the unfair targeting of Black people by police. The accuracy of many facial recognition systems is strongly influenced by a set of training images used to teach the software to identify patterns. Many data sets used for this purpose are predominantly white and male, resulting in biases in the facial recognition system's accuracy. Constructing more diverse training data sets is one approach to solving the problem. Other technical solutions include the development of cyclic generative adversarial networks, disentangled representation networks, and dynamic face matchers. Non-technical solutions include legal restrictions on the use of facial recognition technology by law enforcement. Another non-technical approach involves supporting more diverse engineering teams who would likely be more prepared to take on the challenges of racially biased software. In this report, we advocate for increased diversity in STEM.

INTRODUCTION

When someone takes a driver's license photo, they typically do not expect the picture to end up in a police investigation. However, the driver's license photos of over 117 million Americans have been stored in databases, and law enforcement has the power and technology to use these images to identify individuals (Garvie, 2016). They use facial

recognition systems, a type of artificial intelligence (AI), to find the shapes of human faces in photos, videos, or sketches and to match those shapes to faces in databases of driver's licenses and mugshots. Currently, police do not need a warrant to identify someone using a facial recognition system. In Texas, the Department of Public Safety maintains a facial recognition system with a database of 24 million driver's license photos, and the law states that this system can be used to aid in criminal investigations (Texas Transportation Code, 2015; Center on Privacy & Technology, 2016).

This ability may be considered a breach of privacy even if these systems were perfectly accurate. Unfortunately, the systems are not perfectly accurate. False identifications occur and follow a racially biased pattern. Facial recognition systems are most likely to falsely identify people with dark skin tones--people already at heightened risk of being unfairly targeted by police (Jeffers, 2019).

This report focuses on racial bias in facial recognition systems, the risk posed to the public by the use of these systems in police investigations, and possible solutions. To provide background for understanding facial recognition systems, we will describe how artificial intelligence in general works. The accuracy of these systems depends heavily on the data used to train them. Image datasets used to train facial recognition systems are racially biased. Our report will cover the extent of the bias and the impact on the public before exploring potential solutions. One solution to address biases in facial recognition applications is to construct better training datasets. Another approach to the problem involves implementing legal restrictions on the use of facial recognition technology. We will consider a range of technical and non-technical approaches to mitigating racial bias in facial recognition technology.

HOW ARTIFICIAL INTELLIGENCE WORKS

Artificial Intelligence (AI) is an exciting tool in science and engineering because it has the potential to solve many problems across disciplines (Safdar et al., 2020). Many institutions are incorporating AI software into their operations including police organizations, healthcare, and education (Safdar et al., 2020; Andrejevic & Selwyn, 2020). AI software can automate processes that are usually considered cognitive such as decision making and pattern recognition (Intahchomphoo & Gundersen, 2020). Though these capabilities open up exciting possibilities such as self-driving cars and the ability to test spacecraft without leaving earth's surface, the technology is still developing, and its ethical implications must be analyzed (Intahchomphoo & Gundersen, 2020). The main purpose of this report is to discuss the social impact of facial recognition systems which rely on artificial intelligence software.

To fully understand the consequences of artificial intelligence software, it is important to understand how this software works. AI is a broad category of algorithms that includes supervised and unsupervised machine learning, which themselves, are broad categories. The facial recognition systems discussed in this report utilize both of these techniques. Ideally, this software will take a set of inputs and match each input to a label (IBM Cloud Education, 2020), much like a child with a set of differently shaped blocks might match each block to a slot of the same shape. However, whereas a child learns to visually recognize the shape of a block, an AI algorithm must learn to look for mathematical patterns in the input and associate these patterns with a particular label.

The ability of AI to correctly identify mathematical patterns depends heavily on the system's training data. During development of a supervised machine learning system, an AI system is fed sample inputs which are already labelled (Sandbach et al., 2012). The machine analyzes inputs in this training data set which have the same label and makes inferences about how the inputs are similar. After training, when the machine is given unlabeled inputs, the algorithm should be able to determine whether the input is similar to something the algorithm has seen before and assign a label accordingly (Sandbach et al., 2012). Due to this learning process, flaws in the training set can lead the AI system to draw erroneous conclusions. For example, this author is contributing to a supervised machine learning project which identifies wildlife in images. The software is still in development and classified this author's dog as a whale because it has never seen a picture of a dog before. By contrast, Microsoft Word's image identification software provides the detailed label: "a dog sitting in the grass".

Figure 1: A dog is misidentified as a whale. The red box is meant to locate an animal within the picture. The large red box and the confidence score of 0.41 (out of 1) demonstrate that the computer is unsure of the classification. (Sandbach et al., 2012)

A facial recognition system is developed by training the software on a dataset containing images of faces. For the software to be effective, the algorithm has to be trained to handle many challenging images. For example, the software should be able to recognize faces in the input image even if the faces are partially obscured by their angle, position, or light intensity. The software should also be able to recognize someone regardless of their facial expression. To handle these challenges, once the facial recognition software locates a face, it performs normalization. The normalization process involves determining the relative location of features in the face, and mathematically transforming these features to rotate and rescale the face (Garvie et al., 2016). An example is demonstrated in Figure 2. Once the face is normalized, the facial recognition system can extract additional details to make a comparison to known faces and calculate a similarity score (Garvie et al., 2016). As is the general case in

supervised machine learning, the success of the identification process depends on the quality of the training data. Unrepresentative training data sets can result in racial disparities in the accuracy of facial recognition systems (Intahchomphoo & Gundersen, 2020), a topic that will be further explored in the next section.

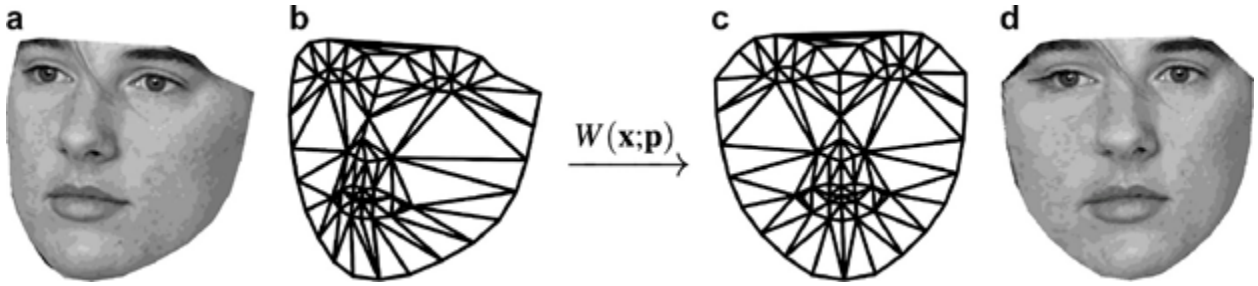

Figure 2: In the normalization process, the features of a face at an angle are mathematically transformed so that the face is facing forward (Haghighata, 2016).

HOW FACIAL RECOGNITION DEMONSTRATES BIAS

Facial recognition is vulnerable to the same pitfalls as general supervised machine learning. The software will demonstrate more error when attempting to identify faces if there were not many similar faces in the training dataset. Many datasets used to train facial recognition systems include disproportionately many light skinned people and males. For instance, white people make up 83.5% of the Labeled Faces in the Wild data set, and light skinned people make up 79.6% of the IJB-A dataset (Buolamwini et al., 2018). These sets are used by many developers as benchmarks to judge the accuracy of facial recognition systems (Buolamwini et al., 2018). Additionally, racial bias in image data long predates modern image classification. Most photography techniques have been calibrated to accurately capture light skin tones. As a result, the software can extract information from images of white people more easily but may find images of black people to be more “challenging” (Kewis, 2019).

The racial biases that go into creating facial recognition systems have been proven to have consequences. In a research study conducted by Klare et al.(2012), the researchers found that the effectiveness of facial recognition systems on certain demographics is dramatically affected by the training sets used to train that system. The facial recognition system is typically tested by making it scan a picture and identify a face that most accurately matches the face from the picture using a certain database. The test is considered to be a misidentification when the face that the system chooses is not the correct face. The researchers found that standard facial recognition systems

are 5 to 10% less accurate when identifying Black people compared to light skinned people, depending on the particular system (Klare et al., 2012). Similarly, they found that these systems are between 5 and 8% less accurate when identifying women (Klare et al., 2012). In their research, Klare et al. also determined that by training the system with a data set containing only black people, they could reduce the accuracy gap to less than 1% (Klare et al., 2012).

This work also demonstrates that the background of the developing team has an impact on software biases. Facial recognition systems developed in the west perform best on white males while facial recognition systems developed in the east perform best on Asian males (Phillips et al., 2011). This comparison suggests that in some cases the demographics of the team impacts the demographics of the data set used to develop the facial recognition system. Additionally, people of color may be more likely to identify racial biases, since they are more likely to be affected by these biases. More diverse teams have also been found to be more innovative, so a diverse workforce would be more likely to find technical solutions to the racial biases in facial recognition systems (Parrotta, 2014). Unfortunately, Black and Hispanic people currently make up only 16% of the STEM workforce even though they make up 27% of the overall U.S. workforce as of 2016 (Funk, 2018). More diverse teams would likely create better performing software, and an increase of representation in STEM is one of the solutions we discuss later in this report. The next section will discuss the negative impacts of racially biased facial recognition systems, and later sections of the report will delve further into solutions for the problem.

HOW BIAS IMPACTS THE PUBLIC

By adopting a critical attitude towards facial recognition technology, engineers and scientists can prevent harm to the public. As engineering students, we believe that promoting ethics in STEM is important for fostering public trust in engineering and to do our professional jobs well. In their work on engineering ethics, Harris et al. (2019) argue that professional care for the well-being of the public is a virtue that engineers should work to achieve. Through this project, we wish to advocate for protecting the public from the harms of biases in image classification technology. Davis (1991) defines the public with respect to engineering as the set of all people vulnerable to engineering decisions due to some barrier to becoming informed about the risk of those decisions. We adopt this definition.

However, it is essential to note that some people who are well informed about the risks of image classification technology do not have the option to protect themselves from those risks. Facial recognition is increasingly used in many contexts, including airport security, public schools, and police investigations (Andrejevic & Selwyn, 2020). Without

legal protection or intervention by engineers, people have limited ability to opt out of the use of facial recognition systems in these contexts. We argue that such people should also be considered members of the public, and engineers have a responsibility to protect them as well.

We believe the Respect for Persons (RP) approach to ethics is an important tool for evaluating the harms the public may face because of facial recognition technology (Harris et al., 2019). The “fundamental principle” of this ethical framework is preserving the autonomy of all humans (Harris et al., 2019, p. 36). To follow this approach, we must recognize the freedom of all people to make decisions for themselves. Each person has the right to freedom, life, and safety from harm. Engineers should avoid contributing to work that violates these rights.

The RP approach to ethics is preferred to other popular frameworks because its emphasis is on protecting all members of the public from harm. By contrast, utilitarianism--which prioritizes an outcome that maximizes overall benefits--may sweep the harm caused to a minority of individuals under the rug (Harris et al., 2019). A virtue ethics approach would focus more on whether a particular person makes good choices than on the consequences of those choices for others (Harris et al., 2019). Through the RP framework, this report seeks to bring particular attention to harms to the public caused by the use of facial recognition systems by police.

When law enforcement officers upload an image to a facial recognition system, the algorithm compares the image to each face in the available database, producing a similarity score (Finklea et al., 2020). In practice, the system returns every image with a similarity score above a certain threshold, rather than just one match. There is no guarantee that a true match is included in the images produced by the facial recognition system, and several of the images produced are false positives. At this point, the task of making a final match falls to a human being (Finklea et al., 2020). Research into the use of facial recognition by passport officers has found that both untrained students and passport officers have the same error rate when choosing the correct face out of a selection provided by the facial recognition system—over 50%. Even highly trained “facial examiners” have a 30% error rate (Dunn et al., 2015). Thus, humans introduce significant error into the identification process which may lead to false arrests. The harms caused by these misidentifications are compounded by the racial biases in facial recognition technology

The previous section reported that facial recognition systems show more error when identifying people of color and the most error when identifying women of color (Garvie et al., 2016; Hardesty, 2018; Albiero et al., 2020). Members of these groups are already marginalized within society, and biased facial recognition can exacerbate this

marginalization. Black people are disproportionately more likely to be subjected to unjust incarceration (Jeffers, 2019). People of color make up 67% of the prison population but only 37% of the U.S. population (Jeffers, 2019). Unjust imprisonment is a clear cut violation of people's autonomy.

Both the biases in facial recognition and the way these systems are used by law enforcement could contribute to the mass incarceration of Black people. The Federal Bureau of Investigation (FBI) divides its image database into civil and criminal images, and local law enforcement agencies use the criminal database to try to find matches for persons of interest (Prest, 2019). The moment someone is convicted of a crime their images are moved into the criminal database, and absent a court order, their images will remain there until 110 years of age, "or seven years after notification of death with biometric confirmation" (Prest). Thus, populations that are already disproportionately targeted by police are likely to gain even more police attention because of this system.

The inaccuracy of facial recognition technology directly contributes to the unjust incarceration of Black people, so it is also directly responsible for creating a cascade of harmful effects. In addition to the threat of incarceration, Black people face decreased professional opportunities if they are investigated by police. The process could also cause severe psychological stress, even to innocent people (Jeffers, 2019). This stress could also contribute to health impairments which increases mortality in the Black community (Juster et al., 2010). Since biased recognition software contributes to harm and restricted autonomy for Black people, it contributes to violations of their rights.

Furthermore, engineers should also be aware that even if facial recognition systems operate perfectly, the software may be a risk to people's privacy. The software is designed to analyze data about individuals and make inferences about their demographics. Thus, unbiased facial recognition systems could be used by biased people to target members of the public based on their background (Castelvecchi, 2020). For example, biased police officers could still use unbiased software to target people based on race. Andrejevic & Selwyn (2020) also express concern that rapid integration of machine learning into society may lead to an increase in automatic decision making based on machine learning outputs. These decisions may overemphasize an individual's race. Thus, even if the software itself is morally neutral, it may be used as a tool to deprive members of the public of their autonomy or otherwise harm them. As part of the RP approach to ethics, engineers should advocate for people's right not to have data collected about them that could be used against them. There is a need both to improve software and to provide protections for the privacy of members of the public.

HOW THE PROBLEM CAN BE APPROACHED NOW

There are both technical and non-technical approaches to this problem. An approach is considered technical if an engineer directly applies skills encompassed by the engineering profession (Harris et al., 2019). Engineers are well positioned to address this problem because their technical background acquired through professional experience gives them insight into how the problem arises. They can use this insight to craft technical solutions and to communicate clearly with the public to advocate for both types of solutions. The act of communication itself can be considered a non-technical approach to the problem as can the other status quo solutions discussed in this section.

Communication is an important part of helping the public form and voice opinions on this issue. Bromberg et al. (2020) conducted research which asked people their opinion of police use of facial recognition systems. The results of this research suggested that over 60% of the US population would voice support for this use of facial recognition systems if asked publicly but support dropped by around 20% when these populations were asked anonymously. The researchers theorize that the study's subjects were reluctant to reveal their skepticism to others because they feared judgement for holding an opinion that was not considered socially acceptable (Bromberg et al., 2020). These results show a need to create room to discuss concerns about facial recognition systems. Through this report, we hope to provide people with the support and information they need to speak up for themselves. Providing people with information is in line with the RP approach to ethics because the process provides people with tools to exercise their own autonomy.

We also advocate for efforts to create a more diverse STEM community as another non-technical solution. Facilitating diversity is an important goal in itself because everyone should have equal opportunities within STEM. As an extra advantage, we note that non-technical approaches to increasing diversity in engineering teams may lead to technical solutions to the problem as these teams may be more likely to construct diverse training data sets which will result in more accurate facial recognition systems.

People of color are underrepresented in STEM for a host of reasons. For example, many schools and workplaces do not put enough emphasis on increasing racial and ethnic diversity. Roughly 57% of Black people in STEM say their workplace pays too little attention to increasing racial and ethnic diversity (Funk, 2018). For comparison, around 43% of Black people outside of STEM say the same of their workplace (Funk, 2018). Therefore, there is a need for a push for workplaces to put more focus on racial and ethnic diversity (Funk, 2018).

The existing education system also perpetuates racial imbalance in STEM. About half of STEM workers believe limited access to quality education to be the major reason for the

lack of diversity in STEM (Funk, 2018). Studies have shown that students who foster relationships with faculty tend to persist to graduation, and the lack of minority faculty members leads to minorities making fewer of these relationships (Cole & Espinoza, 2008). Despite this fact, 48% percent of engineering schools had no Black tenured or tenure-track faculty members as of 2016 (Robinson, 2016).

Another force limiting access to education are institutional racist stereotypes that Black and Hispanic students don't belong in STEM. One study interviewed 38 Black and Hispanic university students and found that this type of racist stereotype is rampant in both minority serving institutions and predominantly white universities (Mcgee, 2016). This type of racist hostile environment leads to Black and Hispanic students experiencing imposter's syndrome; high achieving minorities students cope with these problems by trying to prove themselves and often feel burnt out by their efforts (Mcgee, 2016). A major step towards improving diversity in STEM would be to push for universities to support the careers of faculty members belonging to minority groups and to examine the institutional racism that exists in STEM fields. The benefits of increased representation in STEM would go far beyond developing innovative solutions to facial recognition technology.

Another current non-technical approach to the problem of racial bias in policing involves banning facial recognition technology (Lunter, 2020). Attorneys for the American Civil Liberties Union (ACLU) have called for this course of action on the basis that facial recognition technology is too inaccurate to be put into practice. San Francisco was the first city to take this action and at least 10 other US cities have followed suit (Lunter, 2020; Castelvecchi, 2020). Such bans are important because they formally protect people's rights and autonomy, rather than leaving the power in the hands of engineers, technology, or police. However, this approach would mean we would have to forego the efficiency that comes with facial recognition technology. Facial recognition drastically reduces the time it takes to identify suspects. Plus the absence of facial recognition could even magnify the racial bias problem. As we have already discussed in the passport study, humans are usually even worse at identifying faces of other people, especially people of other races.(Dunn et al., 2015) Without facial recognition technology which could possibly be optimized to reduce its biases, identification of suspects would now fall on humans that have the same biases.

Without laws in place to restrict the use of facial recognition technology, law enforcement officers take advantage of their broad powers. For example, the Federal Bureau of Investigation (FBI) has failed to implement several recommendations made by the Government Accountability Office (GAO) relating to their use of facial recognition systems. The recommendations included timely publication of privacy impact assessments (PIA) of the FBI's facial recognition system and testing to ensure that

identifications made by the system are accurate for all search sizes. The FBI did not comply, stating that the recommendations went beyond what the law requires and that the system's accuracy is sufficient for their investigative purposes (Government Accountability Office, 2016). Thus, even if effective technical solutions are developed in the future, legal protections should be put in place to check the power of law enforcement and protect the rights of the public. A realistic non-technical (policy) solution could be the enforcement of stricter regulations on the accessibility of subject data for governmental bodies, and third party companies that profit from using these technologies. The current solutions outlined here are works in progress, and the coming sections will address their limitations as well as up and coming solutions.

HOW CURRENT SOLUTIONS ARE LIMITED

Although great strides in gender and racial diversity in the workforce have been made in recent years, there still exists a large disparity, especially in STEM professions. Implementing policies that make it more affordable for underprivileged minorities to pursue professional endeavors in science and technology would ultimately make the training datasets for facial recognition systems more gender and racially balanced. Perhaps, if members of the upper class of our society were taxed more often, and at a greater rate, some of these funds could be used by local governments to deliver educational grants and scholarships to minority students.

Although the non-technical solutions presented in this report constitute reasonable long-term solutions, for these solutions to be effective, many other dominoes in our society would need to fall. Put differently, in order to increase minority representation in STEM occupations, greater public awareness must first be spread about the racial stratification in the workforce in today's society and the associated imbalance in training data that is made available to developers of facial recognition systems. In turn, members of underprivileged racial groups must then be given additional opportunities through grants and scholarships to pursue degrees in science and technology. For this to happen on a national level, governmental bodies would need to implement new policies such as taxing the wealthy at greater rates and restructuring their budget in order to fund these opportunities for the underrepresented.

For these reasons, this report will present several technical solutions whose results would likely be noticed in a much quicker timeframe. The source of this difference in time lies in the lesser number of hurdles that facial recognition developers would have to jump through to implement their technologies in law enforcement. In particular, the only barrier that lies in the way of implementing new facial recognition technologies in law enforcement (assuming the technologies abide by ethical codes) would be the

manufacturers and the governmental bodies themselves. By contrast, there exist many more roadblocks to implementing the non-technical (policy) solutions mentioned above.

HOW TO SOLVE THE PROBLEM IN THE FUTURE

The up and coming solutions that we describe here can be considered technical solutions which follow a radical design process. These solutions involve engineers applying their professional expertise to create better facial recognition software. The concept of radical design was first defined by Walter Vincenti (Van Gorp & Van de Pol, 2008). He first defined normal design as a process by which engineers create technology that conforms to established conventions. An example would be the design of a pressure vessel that works the same way as other pressure vessels and is physically similar to them as well (Van Gorp & Van de Pol, 2008). By contrast, radical design requires some decision making which results in a device that is innovative in some way. While many of the technical solutions we describe make use of machine learning techniques that already exist, the application of these techniques to facial recognition systems is new. Researchers at the cutting edge are developing new machine learning technology that mitigates racial biases, and we classify this work as radical.

There are several supervised and unsupervised machine learning techniques that could reduce racial biases in facial recognition systems if implemented on a large scale. In supervised machine learning techniques, models are trained to make inferences using datasets that are pre-labeled, and under human intervention. By contrast, in unsupervised machine learning techniques, models learn to identify patterns in—and draw inferences from—datasets that are not labelled. This process occurs with little to no human intervention. Examples of techniques that will be explored in this section include generative adversarial networks, representational disentangling networks, and dynamic face matcher technology.

First, we discuss adversarial generative networks, a domain of supervised machine learning models. This class of networks comprise two parts: a generative network and a discriminative network. The generative network receives input from the images in the training dataset and uses statistical techniques to generate a new image with the same mathematical properties as the original copy. The discriminative network receives the synthesized image and attempts to classify the image as either the original image or a synthesized image. In turn, the generative network receives feedback from the discriminative network regarding this classification. This process continues until the generator achieves a target accuracy for constructing synthetic images that resemble the original images and until the discriminator reaches a target accuracy for

distinguishing synthetic and true images. Both of these target accuracy values may either be an unchanging parameter given to the algorithm at the beginning of testing or a parameter that is dynamically updated throughout testing.

The second up and coming solution discussed in this report is a facial recognition system based on disentangled representation networks. These networks are a type of unsupervised machine learning technique. This class of unsupervised techniques breaks down (disentangles) individual attributes of an input image into “high” and “low” dimensional encodings. Put differently, the features of the image are broken down into individual variables, and encoded as separate dimensions whose scope may be large or narrow respectively. For example, an algorithm in the model that reads in an image of a businessman may break down features that are large in scope such as articles of clothing (shirt, shorts, shoes, etc.) into high dimensional vectors (encodings) that may be used to accurately predict the sex of the individual. By contrast, an algorithm whose purpose is to classify features that are narrower in scope (such as facial attributes), reads in an image and breaks its features into low-dimensional vectors. These vectors can be used to classify facial features such as eye color, nose shape, jaw structure, and other properties used for facial identification. These encodings are typically represented as integer-based vectors of red-green-blue (RGB) values.

Finally, the third up and coming solution we discuss is a dynamic face matcher. This technology refers to a class of facial recognition systems that allows for users to choose from different types of supervised and unsupervised machine learning models that are clustered together into a single system depending on the application. For example, users may specify that the facial recognition system should use a model designed to have high accuracy when identifying Black people. Additionally, if the users found that training and buying multiple datasets was too expensive, they could opt to use facial recognition algorithms that are trained on an equally distributed (diverse) database. Each of these up and coming solutions is discussed in greater detail below.

PROPOSED TECHNICAL SOLUTION #1: CYCLICGAN

One type of adversarial generative technique that could potentially solve our problem is the cyclic generative adversarial network (CyclicGAN) proposed by Yucer et al. (2020). This network attempts to add consistency and balance to facial recognition on an individual subject level. In this research, a generative adversarial network was used to perform racial transformations on input data to create synthesized images that maintain identity-related features (such as jaw structure, and eye color) while eradicating racially-dependent features (such as skin tone, and nose shape). Then, a convolutional neural network (CNN) was used as the discriminator network to reconstruct the original

image using the properties of the racially-transformed image. CNNs are a type of machine learning algorithm designed to imitate the network of neurons in a brain. Each “neuron” in the network receives input information, performs a computation, and sends the output to other “neurons”. By implementing this synthesis and reconstruction process, the researchers could isolate facial features of the subjects and create a facial recognition model independent of racial compartmentalization.

The research team conducted their study through three phases: data sampling, image-to-image race transformations, and testing of adversarial algorithms. Through this process, the researchers determined that this technique improved facial recognition accuracy on minority subjects in the training data within the range of 0.38% - 1.51% relative to the industry standard facial recognition algorithms that use racial compartmentalization. Although this figure seems marginal, the researcher’s radical design process created a new way of approaching facial recognition development through incorporating algorithms that are racially-independent.

However, incorporation of racially-independent facial recognition algorithms would not come without limitations in use and additional concerns. In particular, though the training data used in this research was demographically diverse, it was compiled from only a few existing data sets. The authors recommended additional testing that uses a greater number of inhomogeneous data sets. This testing may boost the government and public confidence that the technology can accurately identify individuals across racial spectrums.Then, assuming that the results of these associated tests are comparable to those presented in the study, governmental agencies would be more open to funding future projects to make these facial recognition algorithms production-ready.

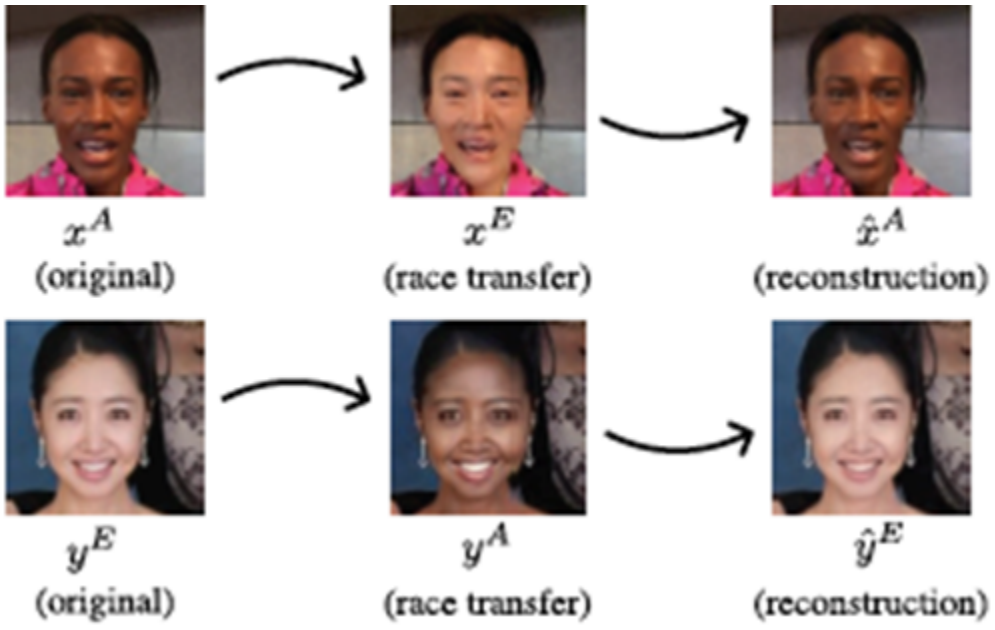

Figure 3: This diagram provides an example of the transformation and reconstruction performed by the CyclicGAN software. This process reduces the impact of racial features in facial identification tasks (Yucer et al., 2020)

PROPOSED TECHNICAL SOLUTION #2: DISENTANGLED REP. NETWORK

An alternative to the aforementioned generative adversarial network approach is the disentangled representation network technique, a type of unsupervised machine learning technique. The research conducted by Xu et al. (2020) addressed the sampling shortcomings that algorithms that perform facial expression recognition—much like facial identification recognition—rely on. In particular, these pitfalls involved an “uneven distribution of subjects in terms of demographic attributes such as race, gender, and age” (Xu et al., 2020, pg.1) in the training models.

In this study, the researchers attempted to mitigate racial biases in expressions such as laughing, crying, and anger, using a disentangled network approach. After running tests using racially diverse datasets with and without data augmentation, the researchers found that accuracy of identification in racial minority groups increased by a margin of between 1.7 - 11.4% from the baseline accuracy levels determined using industry-standard convolutional neural networks (CNNs). Although the improvement of accuracy in identifying facial features of African Americans was reported at the lowest end of this spectrum (1.7%), it still constitutes a foundation for improvement which machine learning professionals could reasonably explore in upcoming years.

PROPOSED TECHNICAL SOLUTION #3: DYNAMIC FACE MATCHER

Klare et al. (2012) proposed the use of a dynamic face matcher to mitigate racial bias in facial recognition software. Facial recognition systems which employ dynamic face matchers would include more than one algorithm for identifying faces. For each facial recognition task, the system would select the algorithm likely to produce the most accurate results. To demonstrate the utility of a dynamic face matcher, the researchers compared the relative accuracy of models trained on different data sets at identifying members of different racial groups. They also compared the accuracy of these models to commercial facial recognition systems, the type which might be used by law enforcement.

The researchers’ objective was to determine whether the demographics of training sets could be altered in order to train models that perform better at identifying members of

particular racial groups. One of their training data sets had an equal distribution of Black, Hispanic, and white faces. Training sets were also created for each individual racial group. The researchers compared the results of algorithms trained on these data sets against three commercially available state of the art facial recognition systems, namely: Cognitec's FaceVACS v8.2, PittPatt v5.2.2, and Neurotechnology's MegaMatcher v3.1. They also compared their results to two facial recognition systems that do not require training on data sets. Instead, these systems use known algorithms to identify faces—the local binary pattern and Gabor features algorithms, respectively. Typically, non-trainable facial recognition systems perform worse at identifying faces than both commercial technology and systems trained by the researchers.

Klare et al. concluded that all commercial facial recognition systems and non-trainable facial recognition systems were less accurate when identifying Black Americans compared to white and Hispanic Americans. However, they also found that if a facial recognition model was trained on only one racial group, then the algorithm's accuracy increased when identifying that group compared to the accuracy of a model that was trained on all groups. For instance, a model trained on the data set of Black people showed a 2% increase in accuracy when identifying Black people compared to a model trained on an equally distributed data set. Likewise, a model trained on only white people showed a 1.5% increase in accuracy when identifying white people. The researchers also tested a model trained on only Hispanic people but did not find conclusive evidence of an increase in accuracy in this case. More testing may reveal the promises and limitations of selective training. The researchers also observed that the model trained on the evenly distributed data set demonstrated a more equal success rate across all races compared to commercial and non trainable facial recognition systems.

Based on these results, the researchers believe that a dynamic face matcher could reduce racial biases in police investigations. Using this system, investigators could choose the best type of facial recognition system for each identification task. Since the dynamic face matcher includes multiple machine learning algorithms, such a system could incorporate the other up and coming solutions addressed in this report. For a particular image identification task, a facial recognition system may choose between a cyclic generative adversarial network, a disentangled representation network, or models trained on specialized data sets. If police argue that training and buying multiple datasets is too expensive, they could still improve on the status quo by opting to use facial recognition models trained on an equally distributed database. This study demonstrates that though the commercial facial recognition systems that are currently available fail to accurately identify Black people, new technology could mitigate the inaccuracies.

HOW THE UP AND COMING SOLUTIONS ARE LIMITED

We must note that though the up and coming solutions discussed here are promising, they do not completely solve the problems we described in this report. The solutions reduce the gap in accuracy at identifying people of color compared to white people, and there is promise that future research in this direction might fully close this gap. However, facial recognition software is currently not 100% accurate for any racial group. As long as this inaccuracy persists, there is the risk that the use of these systems by law enforcement will lead to violations of people's autonomy. Even when the software works perfectly, its ability to gather information about people may be used to violate their rights.

Given that the problem cannot yet be fully solved, engineers may use ethical tests to determine the best available solution to promote. For example, they may apply Michael Davis' Harm Test which advocates for choosing the option that does less harm than any of the alternatives (Harris et al., 2019). This test may be extended to suggest that engineers should choose the combination of options that do the least harm. If possible, part of the best solution may be legislation to protect people's rights. The technical solutions described above also promise to do less harm than the software currently in use. These solutions give engineers and members of the public options to rally behind, creating public pressure on law enforcement agencies and policy makers to improve the quality of their facial recognition systems. We are in the process of creating the awareness and tools to improve the status quo.

CONCLUSION

There are racial biases in the accuracy of facial recognition systems, and law enforcement's use of these systems will cause serious negative consequences if solutions are not put into place. The immediate cause of the inaccuracy that facial recognition systems demonstrate when identifying people of color is the overrepresentation of white people in the data sets used to train this technology. Deeper causes include the calibration of photography to more clearly capture light skinned people and the lack of diversity in engineering teams developing this technology. Additionally, when law enforcement officers use this technology, they are provided several potential matches for the person they are trying to identify and must choose the best match themselves. This process introduces more error into the identification process, increasing the risk that Black people will be unfairly targeted by police. We have discussed a range of potential solutions for these biases. Non-technical solutions

include support to build a more diverse STEM community and putting legal protections in place to prevent law enforcement from misusing facial recognition technology. Several technical solutions are also in development: cyclic generative adversarial networks, disentangled representation networks, and dynamic face matchers. Fully solving this problem will likely take multiple simultaneous solutions. Researchers should continue to search for technical solutions, and engineers should facilitate informed conversations about the issue and its impact on the public.

REFERENCES

Albiero, V., Krishnapriya, K.S., Vangara, K. Zhang, K., King, M.C., & Bowyer, K.W (2020) Analysis of gender inequality in face recognition accuracy. *IEEE Winter Applications of Computer Vision Workshops*. https://arxiv.org/abs/2002.00065

Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: critical questions and concerns. *Learning, Media and Technology*, *45*(2). https://doi.org/10.1080/17439884.2020.1686014

Bromberg, D.E., Charbonneau, E., & Smith, A. (2020). Public support for facial recognition via police body-worn cameras: Findings from a list experiment, *Government Information Quarterly*, *37*(1). 1-7. https://doi.org/10.1016/j.giq.2019.101415.

Buolamwini, J. & Gebru, T. (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81:1–15. Conference on Fairness, Accountability, and Transparency

Castelvecchi, D. (2020, November). Is facial recognition too biased to be let loose? *Nature*. https://www.nature.com/articles/d41586-020-03186-4

Center on Privacy & Technology (2016). Jurisdiction: Texas. *Georgetown Law*. https://www.perpetuallineup.org/jurisdiction/texas

Cole, D., & Espinoza, A. (2008). Examining the academic success of Latino students in science, technology, engineering, and mathematics (STEM) majors. Journal of College Student Development, 49(4).

Davis, M. (1991). Thinking like an engineer: The place of a code of ethics in the practice of a profession. *Philosophy & Public Affairs*, *20*(2), 150-167. https://www.jstor.org/stable/2265293

Finklea, K., Kolker, A.F. , Harris, L.A., & Sargent Jr J. F. (2020), Federal Law Enforcement Use of Facial Recognition Technology, *Congressional Research Service* https://congressional-proquest-com.ezproxy.lib.utexas.edu/congressional/docview/t21.d22.crs-2020-crs-2011779?accountid=7118

Funk, C. and Parker, K. (2018, January). Racial diversity and discrimination in the U.S. STEM workforce*. Pew Research Center's Social & Demographic Trends Project.* https://www.pewresearch.org/social-trends/2018/01/09/blacks-in-stem-jobs-are-especially-concerned-about-diversity-and-discrimination-in-the-workplace/

Garvie, C. Bedoya, A.M. & Frankle, J. (2016, October), The perpetual line-up. Unregulated police face recognition in America. *Center on Privacy and Technology, Georgetown University*. www.perpetuallineup.org

Haghighata, M., Abdel-Mottalebac, M., & Alhalabibc, W. (2016). Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*, *47*(1), 23-34. https://doi.org/10.1016/j.eswa.2015.10.047

Harris, C.E., Pritchard, M.S., James, R.W., Englehardt, E.E., & Rabins, M.J. (2019). Engineers: Professionals for the human good. *Engineering ethics: Concepts and cases* (6th Edition, pp. 1-18). Cengage Learning, Inc.

Harris, C.E., Pritchard, M.S., James, R.W., Englehardt, E.E., & Rabins, M.J. (2019). A practical ethics toolkit. *Engineering ethics: Concepts and cases* (6th Edition, pp. 19-49). Cengage Learning, Inc.

Intahchomphoo, C., & Gundersen, O. E. (2020). Artificial intelligence and race: A systematic review. *Legal Information Management*, *20*(2). https://doi.org/10.1017/s1472669620000183

Jeffers, J. (2019). Justice is not blind: Disproportionate incarceration rate of people of color. *Social Work in Public Health*, *34*(1), 113–121. https://doi.org/10.1080/19371918.2018.1562404

Juster, R.P., McEwen, B.S. & Lupien, S.J. (2010), "Allostatic load biomarkers of chronic stress and impact on health and cognition", *Neuroscience and Biobehavioral Reviews, Vol. 35 No. 1, p. 2e16*

Kewis, S. (2019, April). The racial bias built into photography. The New York Times. https://www.nytimes.com/2019/04/25/lens/sarah-lewis-racial-bias-photography.html

Klare, B., Burge, M., Klontz, J., Vorder Bruegge, R., & Jain, A. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security, 7(6), 1789–1801.* https://doi.org/10.1109/TIFS.2012.2214212

Lunter, J. (2020). Beating the bias in facial recognition technology. *Biometric Technology Today*, *2020*(9), 5-7.

McGee, E. (2016). Devalued Black and Latino racial identities: A by-product of STEM college culture? *American Educational Research Journal, 53(6)*, 1626-1662. Retrieved April 10, 2021, from http://www.jstor.org/stable/44245966

Parrotta, P., Pozzoli, D., & Pytlikova, M. (2014). The nexus between labor diversity and firm's innovation. *Journal of Population Economics, 27(2)*, 303-364. doi:http://dx.doi.org.ezproxy.lib.utexas.edu/10.1007/s00148-013-0491-7

Phillips, P.J., Jiang, F., Narvekar, A., Ayyad, J.& O'Toole, A.J. (2011), An other-race effect for face recognition algorithms, *ACM Transactions on Applied Perception (Perception)*, *8*(2), 1-14.

Prest E. M. (2019), Private impact assessment for the Next Generation Identification-Interstate Photo System, *Federal Bureau of Investigation* https://www.fbi.gov/file-repository/pia-ngi-interstate-photo-system.pdf

Robinson, W. H., McGee, E. O., Bentley, L. C., Houston, S. L., & Botchway, P. K. (2016). Addressing negative racial and gendered experiences that discourage academic careers in engineering. *Computing in Science & Engineering, 18(2)*, 29

Safdar, N.M., Banja, J.D., & Meltzer, C.C. (2020) Ethical considerations in artificial intelligence. *European Journal of Radiology*, *122*, 108768-108768. https://doi.org/10.1016/j.ejrad.2019.108768

Sandbach, G., Zafeiriou, S., Pantic, M., & Yin, L. (2020). Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, *30*(10), 683-697. https://doi.org/10.1016/j.imavis.2012.06.005

Texas Transportation Code, Tex. Stat. § 521.059 (2005 & Supp. 2015) https://statutes.capitol.texas.gov/Docs/TN/htm/TN.521.htm

Van Gorp, A., & Van de Poel, I. (2008). Deciding on ethical issues in engineering design. *Philosophy and design: From engineering to architecture* (pp. 77-89). Springer. https://doi.org/10.1007/978-1-4020-6591-0_6

White D., Dunn J.D., Schmid A.C., & Kemp R.I. (2015) Error rates in users of automatic

face recognition software. *PLoS ONE 10(10)*: e0139827. https://doi.org/10.1371/journal.pone.0139827

Xu, T., White, J., Kalkan, S., & Gunes, H. (2020, August). Investigating bias and fairness in facial expression recognition. In European Conference on Computer Vision (pp. 506-523). Springer, Cham.

Yucer, S., Akcay, S., Al-Moubayed, N., & Breckon, T. P. (2020). Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June. https://doi.org/10.1109/CVPRW50498.2020.00017