

**A Care Ethics Analysis of Facial-Recognition Software Designed to Predict Potential
Criminals**

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Hannah Clark

March 6, 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.

Signed: _____  _____

Approved: _____ Date _____

Benjamin Laugelli, Department of Engineering and Society

Introduction

In May of 2020 a Harrisburg University group of two professors and a Ph.D. student released a press statement announcing the automatic facial recognition software they had created which was designed to predict if someone is likely to become a criminal. The statement release claimed that the software used a neural network model to predict criminality to 80% accuracy with no racial bias (Urland, 2020). But, following the release of this statement, the software designers received major backlash regarding their ability to judge character and whether their research was even scientific (Noorden,2020). This is not the first case of predictive algorithms designed to be used by law enforcement to target individuals. The press release reopened discussions regarding the ethics of law enforcement use of such predictive technology. However, despite much discussion about the use of the software, not much has been discussed regarding the ethics and social responsibility of the three software engineers who designed this technology. By not considering the role of the models' designers, readers are not able to understand and reflect on the ethics that are a part of being a software engineer.

I will analyze the Harrisburg University criminal predictor facial recognition software through the ethical framework of care ethics, a framework that emphasizes the role of relationships and acting with empathy appropriate to the situation, to examine how the three software designers were morally irresponsible in creating this software in today's society. I will do this by outlining the context this software was created in and then examining how the three developers failed at demonstrating appropriate care as a social responsibility, as an attitude and as it relates to power dynamics. I will demonstrate their lack of care by looking at how the designers failed to take adequate actions and make informed decisions in the design of the predictive facial recognition software through statements they have released to the press.

Background

Harrisburg's software was designed by Professors Nathaniel Ashby, and Roozbeh Sadeghain as well as Ph. D. student and former NYPD veteran, Jonathan Korn (Urland, 2020). The press release regarding their research and software claimed their software predicts criminality solely based on a picture of someone's face. This software was released alongside their research which was titled "A deep neural network model to predict criminality using image processing" (Urland, 2020). The software was intended to be released along with their research paper in late 2020 but both were redacted by the school following a letter in June 2020 that was signed by over 1000 technology experts who were enraged by the perpetuation of the "tech to prison pipeline" (Rowe, 2020).

Literature Review

The controversy of predictive policing has existed for many decades and subsequently a vast selection of scholarly sources discuss issues that surround using such technologies. The use of predictive policing methods by different organizations is a common topic of academic debate and scholarly work. Despite the plethora of insight provided regarding the use of predictive softwares, these articles fail to make a judgment of the ethics behind creating a criminal predictive software.

The research article *The Challenges of Prediction: Lessons from Criminal Justice* by David Robinson provides a thorough reflection on the application of predictive softwares by different bodies of government. It begins by providing an overview of the many applications of predictive software by the government and the growth of quantitative data to aid in decision

making. It then expands on the use of criminal predictive softwares in two sectors of government, law enforcement and its application in court systems.

In Robinson's discussion of law enforcement uses for predictive softwares, he argues that when a software is based on historical data, extra scrutiny must be used before this data is applied to current decisions. This is because the implications of these softwares can lead to physical violence and sometimes deadly force (Robinson, 2018). Robinson also discusses how the use of predictive policing often becomes a self fulfilling prophecy. When a software is used by law enforcement and it is shown that one demographic in an area is more likely than another to commit a crime, law enforcement will send more police officers to that area (Robinson, 2018). This causes them to see more crime there, hence validating the softwares initial assumption.

In his discussion of criminal predictive softwares in the court system, Robinson explains how the sentence a person receives can be heavily influenced by predictive algorithms. One example that is discussed is that courts are more likely to rule against a person if they are suspected of recidivism (Robinson, 2018). Recidivism is the likelihood of an individual to commit a crime again and is something that can be extrapolated from criminal predictive softwares. When a court uses one of these softwares, the ruling becomes less about the crime at hand and more about the possibility of their future actions which Robinson cautions against. In this piece although the ethics of the use of this software are discussed in depth, Robinson avoids making any judgment about the people who are responsible for designing the software in the first place.

Similarly, the article *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice* authors Richardson, Schultz, and Crawford discuss the impacts of predictive policing to prevent criminal activity. The article discusses 13

case studies which are currently under investigation for using “dirty” or inaccurate, skewed and systematically biased data (Richardson et al., 2019). These case studies help to highlight the risks of using these softwares and some of their terrible impacts. The authors argue for immense caution to be used when implementing any predictive policing system while once again avoiding any mention of the creators of the predictive policing softwares.

The article by David Robinson examines the use and implications of predictive softwares used by the government -- specifically in law enforcement and court systems. Additionally, the source by Richardson et al. went further in depth on some of the case studies that highlight the problems that arise from using predictive policing softwares. Several scholars have discussed how the use of these softwares can lead to negative consequences for the people that are targeted. However, no consensus has emerged concerning the morality of the actions and decisions made by the designers of criminal predictive softwares. This paper will question the actions of the three designers of the criminal predictive face recognition software and provide a judgment on the decisions they made through the lens of care ethics.

Conceptual Framework

My analysis of the actions of the three software designers draws on the framework of care ethics which allows me to examine the morality of their decisions in creating this technology. Care ethics is an ethical theory developed by Carol Gilligan and Nel Noddings that does not believe in generating morals by learning moral principles (Van de Poel & Rayackers, 2011). Instead, care ethics focuses on relationships and thinking empathetically in order to understand what is morally appropriate to do in each context (Van de Poel & Rayackers, 2011).

This raises the question of what constitutes as a relationship and how do ethical responsibilities change in different relationships? Relationships in care ethics can be defined as the connectedness of people. The care that is appropriate and necessary in each relationship is dependent on the context of the vulnerability and dependence between people (Van de Poel & Rayackers, 2011). Some relationships can be asymmetric, in which case good care requires the person in power to protect the other party. In this case study, the relationship in question exists between the three software engineers and the public, or the people who are affected by its use.

Care ethics states that the role of an engineer is to create positive impacts for all their clients and customers with their designs (Van de Poel & Rayackers, 2011). This is an application of care ethics in engineering which allows for a more holistic approach and focuses on social arrangements rather than individuals in decision making (Van de Poel & Rayackers, 2011). This approach highlights the social responsibility that is embedded within care ethics and outlines that one must act to benefit society (Ganti, 2020).

Care ethics is largely carried out through actions but an equally important component of a relationship is having a compassionate and caring attitude (Van de Poel & Rayackers, 2011). Care as an attitude demonstrates how a person's motivation is indicative of the responsibility and emotional investment they have in a relationship. It is thus crucial to practice a concerned and compassionate attitude in order to provide appropriate care.

In the analysis that follows, I will use care ethics to examine whether the actions and decisions of the three software engineers in creating the criminal predictive software were morally just. I will begin by analyzing their decisions as they relate to care as a social responsibility, with respect to power and as an attitude.

Analysis

The designers of the facial recognition software used to predict criminality failed to show appropriate care in designing their product. They were not morally responsible because they did not show care in the following three ways; as a social responsibility, as it relates to power dynamics and as an attitude. This is apparent through the discussion of their software in their press release announcement. Their decisions in designing the software show their failure to consider how it may impact society, reinforce the existing asymmetric power dynamic and how it demonstrates an inadequate attitude of care. Using the framework of care ethics, the decisions and actions taken by the programmers in creating the facial recognition software will be shown to have provided insufficient care. The following sections will focus on each of the three forms of care to illustrate how the software designers did not meet the needs of the people who are impacted by their software.

Care as a Social responsibility

To understand how the software engineers, Professors Ashby and Sadeghain as well as Ph. D. student Jonathan Korn failed to provide adequate care, it is important to understand how they fell short of their social responsibility of care. The three researchers have a duty to design new technology that will benefit the people it impacts. In this case, this is not only the purchasers of the software but also all the people who are affected by the use of this technology in society. The nature of the software creators' relationship to the public means they have a responsibility for the societal effects of their product in communities. In the case of the facial-recognition software designed to predict criminality, the programmers failed to understand the context of their technology and how it encourages discrimination and perpetuates racism.

The software's blatant perpetuation of racism is better understood by looking at its historical context. The researchers state in their press release that their criminal predictive software "has an 80% accuracy and no racial bias" (Urlans, 2020). The problem with this assertion is that the predictive software is based on existing criminal data which has historically been racist and intertwined with bias. In the United States, a Black person is stopped by police five times as often as a white person (National Association for the Advancement of Colored People [NAACP], 2020). This means if the criminal predictive software is using data based on arrests, the algorithm contains a built in bias towards predicting criminals who are Black Americans. Additionally, if the software is based on data regarding the conviction rate in the United States, this too reflects a racial bias. Today, one of every three Black males born is predicted to serve a prison sentence in comparison to one of every 17 White males who are born (NAACP, 2020). This shows that regardless of which data set is used by the predictive software, the trends in the criminal data depicts society's history of disproportionately incarcerating people of color, especially Black people.

The software designers claim to have no racial bias in their algorithm shows a lack of understanding for the societal context of their software. Regardless of whether the software is modeled on who is arrested or who is convicted, these statistics have society's bias embedded within them. What this software is actually predicting is who's more likely to be stopped and arrested by a police officer or convicted by a court, not who is likely to commit a crime. The software designers did not create a criminal predictor to help the public but rather a product that perpetuates the harmful biases of a historically racist society.

The actions of the programmers also demonstrates a lack of understanding of the harmful stereotypes it relies on. Professor Sadeghian stated that their software "can extract minute

features in an image that are highly predictive of criminality”. Those features are a reflection of past criminals who are. With 32% of the United States being African American while 56% of the prison population is African American, it is important as a software designer to ask if this is data that is morally just to base future decisions off of (NAACP, 2020).

The programmers lack of awareness for the context of their technology could lead to further harmful stereotyping towards a community that have historically faced discrimination. This is harmful to society and creates an environment that is worse for a minority group. To provide adequate care it is the programmers social responsibility to be aware of the context and data on which they base their design as well as its implications for society.

Care and Power

The designers of the criminal predictive facial recognition software failed to provide care in a second way, as it relates to power dynamics. In the case of the criminal predictive software, an asymmetric power dynamic exists between the designers of the technology and the people it affects. It is the responsibility of the programmers in an asymmetric power relationship to protect the vulnerable party with their technology. In this case it is the software designers’ responsibility to protect all of the people affected by the software, but especially marginalized groups. In order to fully understand the context of power dynamics with regards to a criminal predictive software, it is essential to understand who the three designers were creating the technology for.

According to Professor Ashby, the predictive software is a “ tool for criminal prevention and law enforcement” (Urland, 2020). This is important because the criminal justice system and law enforcement have a history of being used as tools to oppress African Americans. Following the end of segregation in the United States, the mass incarceration of African Americans has

been utilized as a political weapon of oppression (Many, 2014). Creating a policing software that claims to be unbiased but is based on data embedded with racism legitimizes the targeted surveillance and arrests of Black Americans. This displays one way that power dynamics were not adequately addressed by the software designers when creating this technology. The software designers did not consider how this software could further exploit a vulnerable population.

Another facet of the asymmetric power dynamics that are perpetuated by the design of this software is that the software encourages violence. Professor Ashby states that the purpose of the software is for “automating identification of potential threats without bias” (Urland, 2020). The automation of such a decision removes the blame from the police officer. A similar criminal predictive software, the “beware system” was used in Fresno, California which gave threat scores to people in the city (Robinson, 2016). When officers arrived at a situation with people who they believed to be higher risk, the officers--believing their lives were at a greater risk--more quickly resorted to force (Robinson, 2016). This type of criminal predictive software legitimizes and preconditions police officers to act more violently.

The beware system is very similar to the targeting that is provided by the criminal predictive software created by the Harrisburg researchers. A software that insinuates violence goes against the fundamental values of care ethics in engineering; to design something that benefits others and protects the vulnerable. Already in society, Black people compose 13.4% of the population but are over 22% of fatal police shootings (NAACP, 2020). To create a software that validates and predisposes the use of violence against a vulnerable group goes against the fundamentals of care especially in an asymmetric power dynamic. It is the duty of the programmers to protect all of society, including minority groups with what they create. But, the

programmers' software instead encourages violence against Black people, like the beware system, and therefore provides inadequate care.

Despite the harm that could result from this software, some may argue that the programmers have a responsibility to protect all of society and a software that can predict criminals increases safety for the majority of the public. It may appear that using this software to protect the greater society outweighs the cost of over policing some communities. I agree that protecting all of society is equally important and that the programmers probably intended to create a software that can correctly predict criminals to help keep the public safe. However, this view fails to consider the validity of these softwares at predicting criminals at all. In 2010, the Human Rights Data Analysis Group used predictive policing to compare drug use to drug related arrests in Oakland, California (Lum & Isaac, 2016). Over the course of the study there were over 200 times more drug related arrests in non-white neighborhoods however, drug use across the city was estimated to be evenly distributed (Lum & Isaac, 2016). This shows the discrepancy between who commits a crime and who is arrested.

The designers of the criminal predictive software in Oakland created a reinforcement bias, as the software predicted criminals who fit a certain description, more police were dispatched in areas where these people reside, which led to more arrests in the area and a self validation of the software. It is clear from this example that a criminal predictive software does not help to keep neighborhoods safer, it instead encourages high scrutiny of specific groups of the population while neglecting to stop crime in other places. Oakland's data shows that either many non-white residence were falsely arrested or many White people were not arrested because police attention was elsewhere. In either case, the responsibility to protect society with this kind

of software in this asymmetric power dynamic was neglected as soon as a criminal predictive software was designed with stereotypes and pre-existing biases at its core.

Care as an attitude

Finally, the programmers failed to have the appropriate attitude required to provide sufficient care. This is important because the programmer's motivation in applying appropriate care is reflective of their investment in their relationship with the people it affects. The programmers' attitudes show their lack of accountability as creators of a criminal predictive software that has the power to influence society. Despite its potential widespread effects, the three programmers failed to show compassion and a caring attitude when making decisions in designing this software.

The first example of the inadequate attitude of the three designers is seen through their lack of attentiveness to the context and implications of the software's subject matter. This becomes apparent in the title of their research paper that came out in conjunction with their software. It is titled, "A deep neural network model to predict criminality using image processing" (Urland, 2020). The word criminality used here immediately underscores their lack of attentiveness to their topic and carries a negative racist connotation. This word has a history of being racially linked. In a response to the criminal predictive software, thousands of opponents of the software signed a letter claiming that using the word "acts as a proxy for race due to racially discriminatory practice in law enforcement and criminal justice" (Coalition for Critical Technology, 2020). This lack of attentiveness is compounded because one of the designers, Korn, is a former NYPD officer who has had ample exposure to learn the implications of the word criminality.

The attitude of software designers illustrates a lack of critical thinking regarding the biases and prejudices that the software's data is based on. The programmers lacked attentiveness when choosing the title of their paper. This attitude of a lack of care is present throughout their design. The absence of adequate education surrounding these topics shows a lack of investment in the feelings of those affected by the software.

The design of the software also demonstrates a lack of empathy towards the people who will be affected by this product. As the group of designers stated in their press release, the software is based solely on a picture of someone's face (Urland, 2020). Making an assumption based on someone's face alone with no context removes all pathos and human reasoning from a decision. Even without a criminal predictive software, the targeting and scrutiny felt by Black Americans due to police violence and shootings has created 50 million more days of poor mental health for Black people in an average year (NAACP, 2020). To remove a police officer's option for compassion and understanding before making a judgment about a person will only worsen the targeting that many Black Americans experience each day. The software design denies predicted criminals a chance for compassion or understanding before a judgment is passed. This reflects the lack of compassion and care that is extended to potential criminals by the three software designers. Given the severe and sometimes fatal consequences that can occur from being considered a potential criminal, it is clear that an appropriate attitude was not used by the programmers in creating this software.

Conclusion

The decisions and actions taken by the three programmers of the criminal predictive software allow for a judgment to be made in regards to the morality of the care they provided

while creating this software. Their choices present important insights into how they provided care as it relates to their social responsibility, power dynamics and as an attitude. Thus, through the lens of care ethics it can be seen that the decisions of the programmers while creating this software were not morally just in each of these three ways.

As software engineers, the creators of the criminal predictor are not simply responsible for producing a useful product but also one that benefits society. Although the societal impacts of an engineer's design cannot always be predicted prior to its release, without using proper care it is impossible to practice responsible engineering. Being cognizant of the role of an engineer and their product in society is the crucial first step to achieving this. Having the right attitude and providing adequate care is one such way for engineers to reflect on whether they are performing ethical engineering.

Word Count: 3774

References

- Coalition for Critical Technology. (2020, June 23). *Abolish the #TechToPrisonPipeline*. Medium.
<https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16>.
- Criminal justice fact sheet*. NAACP. (2020, July 10).
<https://www.naacp.org/criminal-justice-fact-sheet/>.
- Ganti, A. (2020, December 22). *Social responsibility*. Investopedia.
<https://www.investopedia.com/terms/s/socialresponsibility.asp>.
- Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14–19.
<https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- Noorden, R. V. (2020). The ethical questions that haunt facial-recognition research. *Nature*, (587), 354–358. <https://doi.org/https://doi.org/10.1038/d41586-020-03187-3>
- Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYU Law Review*, 94 (192).
<https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson-Schultz-Crawford.pdf>.
- Robinson, D. (2016, January 14). *Buyer beware: A hard look at police 'threat scores.'* Medium.
<https://medium.com/equal-future/buyer-beware-a-hard-look-at-police-threat-scores-961f73b88b10>.

Robinson, D. G. (2018). The challenges of prediction: Lessons from criminal justice. *A Journal of Law and Policy for the Information Society*, 14(2), 151–186.

Rowe, N. (2020, June). *Experts denounce racial bias of crime-predictive facial-recognition AI*. Morioh. <https://morioh.com/p/fb1c1a07ee0a>.

Travis, J., Western, B., & Redburn, F. S. (2014). The underlying causes of rising incarceration: Crime, politics, and social change. In B. Western, J. Travis, & F. S. Redburn (Eds.), *The growth of incarceration in the United States: Exploring causes and consequences* (pp. 104–129). essay. <https://www.nap.edu/read/18613/chapter/6#129>.

Urland, K. (2020, May 6). *Harrisburg University develops facial recognition software to PREDICT criminality*. <https://www.abc27.com/news/local/harrisburg/harrisburg-university-develops-facial-recognition-software-to-predict-criminality/>.

Van de Poel, I., & Rayackers, L. (2011). *Ethics, technology and engineering: An introduction*. Hoboken, NJ: Blackwell Publishing Ltd. (Van de Poel & Rayackers, 2011)