


Spring June 7th, 2019

Computer Vision Machine Learning and Future-Oriented Ethics

Abagayle Lee Blank
Seattle Pacific University

Follow this and additional works at: <https://digitalcommons.spu.edu/honorsprojects>

 Part of the [Applied Ethics Commons](#), [Business Law, Public Responsibility, and Ethics Commons](#), [Digital Humanities Commons](#), [Other Computer Engineering Commons](#), [Other Engineering Commons](#), and the [Risk Analysis Commons](#)

Recommended Citation

Blank, Abagayle Lee, "Computer Vision Machine Learning and Future-Oriented Ethics" (2019). *Honors Projects*. 107.
<https://digitalcommons.spu.edu/honorsprojects/107>

This Honors Project is brought to you for free and open access by the University Scholars at Digital Commons @ SPU. It has been accepted for inclusion in Honors Projects by an authorized administrator of Digital Commons @ SPU.

COMPUTER VISION MACHINE LEARNING AND
FUTURE-ORIENTED ETHICS

by

ABAGAYLE LEE BLANK

FACULTY ADVISOR, ELAINE WELTZ
SECOND READER, DR. CARLOS ARIAS

A project submitted in the partial fulfillment
of the requirements of the University Scholars Honors Program

Seattle Pacific University

2019

Approved _____

Date _____

Computer Vision Machine Learning and Future-Oriented Ethics (June 2019)

Primary Abagayle L. Blank, Faculty Advisor Professor Elaine Weltz, Second Reader Dr. Carlos Arias

Abstract – Computer Vision Machine Learning (CVML) in the application of facial recognition is currently being researched, developed, and deployed across the world. It is of interest to governments, technology companies, and consumers. However, fundamental issues remain related to human rights, error rates, and bias. These issues have the potential to create societal backlash towards the technology which could limit its benefits as well as harm people in the process. To develop facial recognition technology that will be beneficial to society in and beyond the next decade, society must put ethics at the forefront. Drawing on AI4People’s adaption of bioethics for AI, Luciano Floridi’s distributed morality framework, Kate Crawford’s definition of harms of representation, and Microsoft’s leadership in facial recognition ethics within the industry, this paper explores stakeholder responsibility within CVML to create the best integration of CVML for society. The paper attempts to connect ethics with praxis in making decisions related to CVML.

Index Terms – Artificial Intelligence, bias, Computer Vision Machine Learning, distributed responsibility, error rates, ethics

I. INTRODUCTION

ARTIFICIAL Intelligence (AI) encapsulates Amazon’s Alexa, *Terminator*-type robots, new techniques in facial recognition, and an unknown number of future innovations. This broad spectrum of applications is hard to pin down in a single definition. However, M. Taddeo and L. Floridi, in their article *How AI Can Be a Force for Good*, identify the critical aspects of AI that make it different from past innovations. They define Artificial Intelligence as “a growing resource of interactive, autonomous, self-learning agency, which enables computational artifacts to perform tasks that otherwise would require human intelligence to be executed successfully” [1]. This definition shows that there is something *new* about the nature of AI from previous technological advances. AI is something that can emulate and eventually challenge human intelligence. More than that, AI *learns*.

Society is applying AI in every area of life. This paper, however, will only attempt to tackle one form of AI – that of computer vision machine learning in the application of facial recognition. Computer Vision Machine Learning (CVML) is a specific set of techniques for classifying, recognizing, and interpreting image and video data. CVML is applied in areas as varied as facial recognition, driverless cars, and drone flight. The computer vision part is the machine’s ability to detect an image and “see” what it is looking at. For instance, this part of the process may detect shapes, colors, or contrast in a photo and draw out certain features.

The machine learning part is the discovery part of the algorithm that deduces what an image is featuring based on data that is fed

to it by the developers or its environment. This can either be done in a “supervised” or “unsupervised” way [2]. In a supervised setup, the machine learning algorithm is given photos with specific labels, like “male” and “female,” and the machine then learns that photographs with particular features have certain labels. In an unsupervised situation, the algorithm is given a group of photos and told to build self-made groups based on what it sees the differences are. This may result in a group of male and a group of female photos in the end as well. For facial recognition, both techniques are used in different parts of the process depending on the application.

CVML is a subset of AI that is of key significance in the new AI “arms race.” There are enormous economic and hegemonic incentives for nations to develop the best algorithms as fast as they can, and that pressure leads to deploying these technologies quickly as well. Simultaneously, China, the United States, and several other nations are competing to create the best CVML algorithms. CVML in facial recognition is not a concern for the far future; it is currently in development and various stages of deployment. However, it has not been entrusted with many significant decisions yet, especially in the United States. The incentives of development will lead to deployment soon, and it is essential that there is time for ethical reflection before these systems are complete.

Contrary to popular opinion, the long-term success of AI in general and CVML in particular will depend less on the number of products that can be created using CVML, but on how societies choose to develop and integrate them into their culture. Ethics will have a significant stake in the success of CVML. The ultimate leader of the AI race will be the society that can successfully integrate AI for the public good without facing societal backlash from misuse. This paper looks to explore how the members of a society can utilize an adapted bioethics framework to develop and deploy CVML in a worthwhile and endurable way. This will entail analyzing the risks associated with CVML in facial recognition and looking at how to combat them within each level of society. At the end of this paper, it should be clear how ethics can affect praxis and how it will undoubtedly shape the future of AI.

II. UNIQUE CHALLENGES WITH ARTIFICIAL INTELLIGENCE

In general, AI has many unique challenges regarding ethics. In terms of ethics, there are many ways to approach it. This paper will look at ethics through the lens of bioethics. Bioethics has several principles that are meaningful to AI which will be explained in detail later. While there are many great ways to view ethics, this paper uses a mostly consequentialist approach because the paper is concerned with the effects of specific stakeholder actions and how those impact AI’s future. Before digging into those actions, it is essential to discuss how the conversation

around ethics and AI brings up new issues because of the invasive and expedient progress of AI.

One of the catalysts of these unique issues is the speed of innovation with AI. In the past, technology and labor revolutions took many years. However, every day, technology companies and researchers are discovering and inventing new forms of AI and applications. This is a problem because it has left little time for reflection [3]. People have been neglecting the ethical conversations that need to happen before deployment because of the incentives to move fast within the market and the world stage. New applications of AI sound useful and exciting, but they often end up having unintended consequences.

A major concern related to AI is the extensive amount of data that is required to train AI algorithms. For these algorithms to “learn,” they must study immense amounts of data. In fact, “AI is fueled by data;” therefore, it “faces ethical challenges related to data governance, including consent, ownership, and privacy” [1]. However, regardless of AI’s use of data, data security and privacy are already controversial issues. AI “exacerbate[s]” these challenges, but it does not create them. It is AI’s “autonomous and selflearning agency” that raises its unique ethical challenges, not its requirement for data [1]. The fact that AI is given agency and the ability to learn from human-supplied data is the most concerning part.

Another issue related to AI is what responsibilities it should be given and who should be responsible for the decisions it makes [1]. Do we give it responsibility for targeting in a weapons system? Hiring practices? Loan checks? When do we have enough certainty to give it responsibility? How should it be held accountable? Moreover, if we give it responsibility, can we lose the ability to supervise and our ability to redress errors or harms [4]? These questions need answers before any high-risk decisions are allowed to be made by AI. However, AI has already been given the responsibility to make these types of decisions in several domains.

An additional part of the data conversation is about the data that is chosen to train these algorithms. Should we use data that is reflective of the world we live in? Or data that is representative of the world we *want* to live in? Both of these perspectives code a certain bias into the algorithm – either the bias of the current dominant culture, or the bias of a programmer’s individual values. For example, Amazon recently created an AI to aid in the hiring process. Because the data it trained on was from the company’s real past hiring practices, where it hired mostly men and men were mostly promoted, the AI also hired mostly men, but to an even higher degree [5]. The resumes of successful people at Amazon that were fed into the algorithm were of men, which caused the resumes that were desired by the algorithm to sound similar to the past resumes. The algorithm even learned to penalize resumes that referenced the word “women’s”. For example, resumes that referenced roles like “women’s chess club captain” were viewed as less ideal by the algorithm [5]. As illustrated by this example, AI tends to amplify biases that are already a part of our world. Thankfully, Amazon has decided to scrap this algorithm, but it is unknown how many of these algorithms exist that have not been audited.

What happens if society and tech companies cannot find a way to solve these issues with AI in a way that encourages fairness and serves the common good? Many researchers think there will be an AI backlash that may limit the impact that AI could do for the good of society. If consumer confidence in the safety and stability

of AI is down, firm and restrictive regulations may be implemented that frustrate the efforts of AI innovation. Scientists at the University College London think that “should serious accidents occur or processes become out of control... [AI] could lead to societal backlash... not dissimilar to that seen with genetically modified food” [6]. In the wake of the GMO backlash, government entities placed restrictions that some scientists in the field think are unfounded and limit the benefits of GMOs. Further, consumers have lost trust in GMOs and avoid their consumption.

M. Taddeo and L. Floridi think that ethical forethought and regulation surrounding AI “is a complex but necessary task” because the alternative may lead to “rejection of AI-based innovation” and “a missed opportunity to use AI to improve individual wellbeing and social welfare” [1]. Things like “fear, ignorance, misplaced concerns or excessive reaction may lead a society to underuse AI technologies below their full potential... for the wrong reasons” [4]. Like with GMOs, Taddeo and Floridi think humanity made a similar blunder during the industrial revolution by not foreseeing its impact on labor forces and the environment [1]. In order to recover from the industrial revolution and to protect against human rights abuses, there have been many hard-fought struggles. Those could have been less necessary if the industrial revolution was overseen and mitigated ethically from the beginning.

The most obvious fear is wilful misuse of AI, be it for greed, geopolitics, or malicious reasons [4]. Current ills of society may be intensified, and others may be created with the help of AI. This may then encourage the underuse of AI in other sectors. Even with entirely good intentions, tech companies are already facing the unintended consequences of their actions. With no way to assign blame or have mechanisms for reparations, society as a whole may decide AI is no longer worth it. Whether it is fear of “overuse or misuse” [4], cultures may decide to rely less on AI and miss important things it can do.

To avoid this underutilization, ethics must be incorporated into AI at the beginning with governments basing regulations on ethics, technology companies creating standards and best practices, society having correct assumptions about the future of AI, and programmers implementing ethically based algorithms. In the paper *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, the authors discuss how this ethical approach to AI creates a “dual advantage” [4]. The first advantage is for organizations to “take advantage of the social value that AI enables” [4]. This advantage manifests itself in “being able to identify and leverage new opportunities that are socially acceptable or preferable” [4]. Companies that take the ethical approach can use AI in ways that society needs and will appreciate the most. The second advantage is for organizations “to anticipate and avoid or at least minimise costly mistakes” [4]. They can avoid situations that, even if legally unquestionable, will be socially unacceptable and also begin mitigation if there are unavoidable risks [4].

The benefits of ethics in decision making for organizations in the realm of AI should be obvious, but the issue of what framework to start from has been up for discussion. One suggestion that has been proposed, again from *AI4People* and Luciano Floridi, is adapting the already developed framework of bioethics to AI ethics. This gives new application to certain principles already in society’s ethics vocabulary and a rich ethical literature to pull from. The four principles that the bioethics framework uses are Beneficence, Non-maleficence, Autonomy,

and Justice. AI4People suggests that AI requires an additional fifth principle of Explicability [4]. These five principles can guide the solutions and risk assessment of new AI technologies and help determine the best courses of action. They also can inform government regulation and court precedent.

Beneficence is the most straightforward principle. It is all about creating benefits for society through the medium of AI. This category includes things like general well-being, human dignity, and helping the planet [4].

Non-maleficence is also easily applied to AI. In simple terms, this means “do no harm” [4]. AI technology should not be intended for harm and not easily twistable to harm. One of the most significant applications of this principle for AI and specifically CVML is personal privacy. Any AI technology developed under this ethical framework should avoid the infringement of privacy as well as maintain the security of personal data. Creators should think through the capabilities that a new technology can bring into society and determine the risks associated with those capabilities, regardless of how the technology is intended.

Autonomy in the bioethics context is the ability to have control over one’s own body and make decisions about health care. In the AI context, Autonomy is related to decision making as well. In this domain, as a society, we have to “strike a balance between what decision making power we give over to AI and what we keep for ourselves” [4]. This means that decisions that involve outcomes that affect people need to always have some element of human oversight. Along those lines, society must maintain the ability to take back decision making power from AI, even after it gives it over.

Justice, the last principle from bioethics, is the ultimate goal of the previous principles. This is about applying AI in situations and having outcomes that are fair and helpful to everyone, not just select groups of people. AI should be helping eliminate problems like discrimination and bias, not creating more of these problems. It also should be working on solving past harms and not creating new ones. For AI, this means that systems should be not just reflecting humanity right now but improving it for the future.

The principle that AI4People added to this list is *Explicability*. Explicability is a combination of intelligibility and accountability. The reason why this is necessary for AI and not bioethics is because AI is often challenging to understand and locked away in proprietary algorithms and hidden systems, whereas in biological contexts, what happens to a person or organisms body is often plain to see and feel. Only a small percentage of people are developing AI, in a small percentage of countries, which means that society has to focus on how it holds these people accountable. This principle is how AI is linked to bioethics and can utilize the framework effectively, as Explicability is necessary to develop the other four principles [4]. AI algorithms need to be able to be understood by society for them to be held accountable, and accountability mechanisms must be created in order for people to be held responsible.

Overall, these five principles are vital in developing and enforcing AI long term and will be referenced throughout this paper as different aspects of CVML and society are analyzed.

III. COMPUTER VISION MACHINE LEARNING ISSUES

The expansion of bioethics into AI is a great place to start when deciding how to regulate and develop AI. The rest of this paper

will look at how to specifically address issues related to CVML which will allow for a more detailed discussion about what roles each facet of society has in building an ethical AI culture. The issues mentioned above are relevant to the entire AI discussion, but CVML has its own specific issues that need to be addressed for its successful development and deployment. The three main issues that CVML deals with are human rights, error rates, and bias.

A. Human Rights

Even if something is possible, it may not be a good idea, such as mass surveillance on a country-wide scale utilizing facial recognition. According to Brad Smith, representing Microsoft, facial recognition inherently “raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression” [7]. Facial recognition is a technology that fundamentally deals with a person’s identity and how identity is recognized and used.

The issues associated with human rights and facial recognition are broad and nebulous. They can be hard to determine because defining privacy violations and civil rights violations vary between nations and cultures.

It is hard to predict some ways privacy can be violated, but some ways seem like blatant abuses of civil rights that may be easier for society to recognize. For instance, facial recognition could give any government the ability to “enable continuous surveillance of specific individuals. It could follow anyone anywhere, or for that matter, everyone everywhere. It could do this at any time or even all the time. This use of facial recognition technology could unleash mass surveillance on an unprecedented scale” [8]. This fear is getting to be more and more possible and is a genuine fear of Microsoft’s Smith. Already countries have implemented systems working towards this goal on minority communities such as China using a combination of facial recognition and GPS tracking to spy on 2.6 million Muslims in the Xinjiang province [9]. China has come under fire for this practice from other nations, but this should give citizens of other countries good reason to ensure their government cannot do the same to them.

B. Error Rates

What makes the human rights discussion even more complicated is that facial recognition technology still has a long way to go. It is advancing at a steady pace and has become very accurate in many cases, but it still frequently makes mistakes [7]. Moreover, even if it has a high enough success rate to justify deployment, there will *always* be an error rate. Figuring out how to deal with false positives is an essential step in ethically deploying facial recognition technology. Questions arise about what will happen if these systems are used and misidentify or classify someone as a criminal. Currently, courts around the world do not have adequate resources to find anyone “responsible” or give anyone moral recourse since there is no precedent for prosecuting an autonomous learning machine [4]. There needs to be an ethical framework that can begin to frame the discussion around facial recognition that can assign responsibility and give society confidence that wrongs will be righted.

Facial recognition systems have improved drastically in the last five years, primarily due to significant innovations in CVML with deep learning [10]. A report published in 2018 by the National Institute of Standards and Technology (NIST) under the U.S.

Department of Commerce showed that algorithms are up to 20 times better than they were a few years ago in searching databases and finding matches [10]. This report tested 127 algorithms by 45 different vendors, a spread the report claims represents the industry. The test was entirely voluntary, so it left out many major facial recognition players. Microsoft participated, but Google, Face++, Amazon, and IBM did not. However, Microsoft was among the highest scoring for accuracy.

Testing involved a database with over 12 million individuals represented. The demographic makeup was not disclosed. The images included were from a set of law enforcement mugshot images, poor-quality webcam images, frames from surveillance videos, and “wild images” gathered from photographers [10]. The most accurate algorithm had an only 0.2 percent error rate on the clearest dataset, whereas in 2014 there was a 4 percent error rate and in 2010 a 5 percent error rate [10].

The report is important to this discussion for several reasons. First, it shows that there are still error rates, even if they are shrinking. Realistically, no matter how good an algorithm gets, there will always be error rates. The stages of testing using clear photographs with well-lit environments were able to achieve less than 1% error rates; however, in the real world, where these algorithms would be deployed, they would likely perform with significantly higher error rates. Even in this report, there was a clear drop off in success when the test set was changed from mugshots to “wild” photos and frames from videos [10].

Second, it shows that there are clear winners and losers. From reading through the report, it appeared that large organizations with access to large datasets, like Microsoft, IDEMIA (A French security and identity company with 3 billion dollars in annual revenue), and Yuti (a Chinese company with government resources) were the most accurate. Others, mostly smaller companies with less access to datasets, performed poorly, sometimes with around 50% accuracy on difficult parts of the test. This is telling because it shows how vital access to data is for the training process and that more resources do often produce better algorithms.

A third takeaway is that this error rate is for *this dataset* – it does not predict how the algorithm will do in any real-world situation or even on a different dataset. Datasets are in themselves inherently not a representation of the real world and often do not predict the accuracy of an algorithm used in the real world. This test is likely an indication of what algorithms are the most accurate in general, but it should not represent the official “error rate” of an algorithm, as the dataset is not likely the same as its practical use.

For instance, Amazon’s Rekognition facial recognition algorithm was tested independently by the ACLU in 2018 and had some alarming results [11]. This system is available to the public and the test that the ACLU performed only cost them around 12 dollars. At the time, ICE and other government agencies were considering using Amazon’s facial recognition resources and the FBI was under contract with them (not to say they were using this exact product). The ACLU ran members of Congress through a mugshot database with 25,000 public images. Out of the members of Congress in both the House and Senate, the algorithm flagged 28 individuals as “criminals.” Both Democrats and Republicans were flagged, young and old, male and female. However, people of color were disproportionately flagged as criminals. 39% of the false matches were of people of color, even though only 20% of the Congress members were people of color. The ACLU was

concerned about the error rate in general. However, they were more concerned about how the error rate affected different ethnicities disproportionately [11].

This example shows how systems approved and tested in specific scenarios can perform poorly with shocking results in a real situation, outside of its training data. More importantly, it highlights how error rates can have a profound impact on real people if society and the algorithm’s creators think that these systems are fool proof. For instance, how much should we trust facial recognition systems to make decisions in criminal trials? For ICE investigations and border security? In the real world, there are real consequences to an error, something that, if treated incorrectly, could send someone innocent to prison or deport them from a country.

China has already publicly experienced the consequences of a false positive, as they have deployed a facial recognition system to catch criminals and jaywalkers on their streets [12]. The government has installed the systems in major cities like Beijing, Shanghai, and Shenzhen. They have been used to identify tens of thousands of jaywalkers and are primarily used as a way to publicly shame those who jaywalk, naming them in a list with their picture and name. The error rate of this system is unknown, but it is not absent, as Dong Mingzhu experienced [12].

Dong Mingzhu is a successful businesswoman in China, a president of China’s top air-conditioning company. One day, her name and photo appeared on a list of jaywalkers in an area she had not traveled through. Later, it was discovered that her face was on a bus in the intersection, and the system analyzed that image and flagged her name. Chinese officials claim that the system has been upgraded to avoid those instances in the future, but if top technology companies have still substantial error rates on clear images in the United States, it is hard to imagine China has a perfect system using real-time video feeds [12]. In fact, Face++, the technology behind China’s facial recognition, has a self-declared accuracy rate of 97.67% [13]. This error rate has not been confirmed by outside auditors, yet this still indicates an imperfect system.

Right now, the consequences of being caught jaywalking in China are the person’s name added to a public list, but they may increase to fines in the future. Additionally, as this technology is used to solve more crimes and in other domains, more consequences could be in store for those recognized. It does not seem like China has any systems in place for recourse or to confirm the results before going forward with the data at the moment, but it seems like that would be wise.

Another example of false positives in the real world is the UK’s trial deployment of automated facial recognition to identify criminals and people on various watch lists in public spaces. At the time of a report written by Big Brother Watch in the UK in 2018, the technology had been used by Leicestershire Police, Metropolitan Police and South Wales Police at several public events. In that time frame, the individual forces were reporting between 91% and 98% *error* rates with thousands of false positive matches and only a small percent of accurate matches. Not all of these false positives were acted upon, especially since several of the reported false positives were obviously wrong with women being identified as men. However, at least twice as many innocent people than those who were actually arrested in one of these deployments by South Wales Police were stopped and forced to prove their identity [14].

The UK’s false positive rate is astronomically high. Civil liberties groups have a right to be concerned about how this is being used, not just because of the error rates but also for human rights reasons. However, consequences for identification are minimal compared to other potential situations. Ultimately, these technologies could be used in some of the most important investigations modern society is undertaking, such as terrorism investigations. The stakes are high to find the perpetrators, but the stakes are also high for a false positive. How do we weigh the risks associated with saving lives and the uncertainty of facial recognition in these situations?

Sri Lanka has had to deal with question firsthand and can serve as an example of what can happen if the balance is too far in one direction. Sri Lankan officials recently misidentified a student at Brown University as one of the Sri Lankan Easter terrorists using facial recognition technology. Amara K. Majeed’s photo was misidentified under the name of the real suspected terrorist and sent out in an alert that was included in several broadcasts. She woke up to 35 missed calls and her social media pages filled with death threats. This false positive not only prevented the real terrorist’s picture from being circulated but endangered many of her family members still living in Sri Lanka and in the States. Not to mention, it was deeply troubling for her to go through. With terrorist investigations, the luxury of taking time to confirm identities is often not present; however, not taking the time to double check results can put even more people’s lives in danger and ruin someone’s life [15].

The examples of Amazon’s Rekognition system, China’s jaywalking system, the UKs facial recognition failings, and Sri Lanka’s facial recognition incident show that currently deployed technologies have error rates and can inflict terrible consequences on undeserving individuals. This is not going to change, even if error rates continue to get smaller. The government and the other stakeholders involved need to decide how error rates are going to be handled ethically, because, in some instances, a false positive could mean the death penalty.

C. Bias

One of the most significant issues facing facial recognition is bias. A recent study at MIT showed that top technology companies including Face++, IBM, and Microsoft had both racist and sexist algorithms, performing with a 34.7% maximum error rate on women of color and 0.8% on white men [16]. Not to leave out Google; the company in 2015 came under controversy for classifying African American faces as gorillas in their Google Photos app [17]. Since then, the service has disabled search results for “gorilla,” “chimp,” “chimpanzee,” and “monkey” with no sign of a real incoming fix [17]. These examples show two things; they show that facial recognition and classification are far from perfect and that the imperfections are often affecting those already marginalized in society the most.

This type of extreme bias is not a problem with CVML specifically; it is an inevitable one that comes from cultures and their own biases. Solving this problem of CVML is going to take more thought than the previous two issues, as it addresses the depths of the human condition. However, using the ethics framework, several potential solutions are available that will likely reduce bias and make facial recognition better for everyone.

In CVML literature, bias is sometimes not about discrimination or “human bias” – often, it means a particular type of bias found

in an algorithm that is statistical in nature and definition. However, what this paper is talking about is the type of bias that the layperson would think of - bias that would indicate that the algorithm is racist, sexist, or ageist or basing its results on stereotypes or wrong assumptions. Kate Crawford, a leading AI researcher and co-director of the AI Now Institute at NYU, says that this bias “is a skew that produces a kind of harm” [18].

Bias is a large part of the potential backlash that AI faces. It has been a significant part of the bad press that AI has received over the past several years. Crawford argues that if CVML systems “keep producing biased results... then people will no longer trust these tools or want to fund this type of work” [18]. In addition to consumers who reject AI, those in the industry building these systems may not want to participate in the process with repeated issues of bias. This increasing climate means that companies and governments will have to prove that their facial recognition technology is unbiased before anyone is going to trust them to use it.

Kate Crawford separates the harms of bias into two categories – harms of allocation and harms of representation [18]. Most of the literature has focused on harms of allocation and includes Amazon’s hiring algorithm [5]. Amazon’s algorithm creates an unfair distribution of resources and opportunities because of a bias in the algorithm. Harms in this category often have economic consequences like who will get approved for a loan or who receives a job offer. Harms of representation are less discussed in the literature, but they are the most relevant to facial recognition. These harms occur when “systems reinforce the subordination of some groups along the lines of identity” [18]. This harm is shown in Google’s algorithm classifying African Americans as “gorillas” [17]. Crawford believes that representation is a long-term problem that needs to be addressed, while allocation is more short term [18]. Representation is often the first step in the chain that leads to unfair distribution of resources and opportunities, as it builds certain views on classes of identities [18]. However, even if no negative results occur because of the representation bias, it is still problematic in itself because of how it treats identity.

Allocation is an immediate threat and it gets attention because of the quantitative impacts it causes. However, representation will become more and more of an issue as stereotypes and assumptions about identities are coded into our CVML algorithms. Allocation is also easier to tackle because it can be quantified. In Amazon’s case, it was easy to tell just how many women were hired and how many were not. Representation takes on cultural and social value and is often hard to detect and formalize [18]. However, often, representational harms are the root of allocation harms.

There are five main ways that representational harms exist (see Fig. 1). The first is *stereotyping*. An example of stereotyping would be associating certain words with specific subclasses like in the paper *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings* [19]. This can also be seen by Google Translate making sexist translations from gender-neutral languages. In one documented case, the phrase “He is a nurse, she is a doctor” translated to and from Turkish (a gender-neutral language) turns into “She is a nurse, he is a doctor” [18]. Google has recently built in mechanisms to mitigate this.

Another representational harm is *recognition*. Recognition harm is evident in facial recognition when those systems do not recognize specific classes of people. Kate Crawford describes this as “failing to recognize someone’s humanity” [18]. This harm is complicated by the fact that different skin tones are more

challenging to recognize from a technological standpoint. As mentioned earlier, many large-scale facial recognition algorithms operated by Microsoft, IBM, and Face++ have a difficult time recognizing women with dark skin in comparison to men with lighter skin [16]. However, technological influences do not seem to account for the wide margin of error between the two, especially since after the study was released the algorithms were improved relatively quickly.

towards others since the beginning of humanity. However, what makes these harms so important when discussing CVML is the perceived neutrality of technology. A problem with bias in our computers and algorithms is that society tends to inherently trust them to give “objective” results, and rarely stops to think that they could be biased. Jaron Lanier, author of *You Are Not a Gadget: A Manifesto*, says that “people will accept ideas presented in technological form that would be abhorrent in any other form”

	denigration	stereotype	recognition	under-representation	ex-nomination
Image search for 'CEO' yields all white men on first page of results.			x	x	x
Google Photo mislabels black people as 'gorillas'	x				
YouTube speech-to-text does not recognize women's voices			x		x
HP Cameras' facial recognition unable to recognize Asian people's faces			x	x	x
Amazon labels LGBTQ literature as 'adult content' and removes sales rankings		x	x		x
Word embeddings contain implicit biases [Bolukbasi et al.]	x	x	x	x	x
Searches for African American-sounding names yield ads for criminal background checks [Sweeney]	x	x		x	

Fig. 1. Harms of Representation [18]

Denigration harms are also evident in facial recognition applications. These harms are realized when technology associates culturally disparaging terms or actions with a person's identity. Google's “gorilla” issue is more than just misidentification; it is also denigration because of the historical use of the word against African populations [17]. This issue is challenging to solve because it often needs human interaction to recognize the disparaging associations and for those to be recognized as harmful.

Underrepresentation is a harm that is found mainly in facial recognition training sets and is often the cause of many of the other harms. Because people of color and other minorities are often missing from search results on the internet, and therefore not entered into datasets, they are often not represented highly in datasets that facial recognition algorithms are trained on. It takes specific and willful creation of training sets to be representative of populations to fix this harm.

Ex-nomination is the harm of eliminating social identity by almost ignoring its existence. This term comes from Barthes where he coined it to describe what the bourgeoisie do to hide their name and identity by not referring to themselves as such to naturalize bourgeois ideology [20]. This can show up in some of the same examples as mentioned above, as ex-nomination can present itself in technology not recognizing a certain class of people with facial recognition technology or by having implicit biases towards certain adjectives to describe certain classes [16] [19].

Many of these harms have examples outside of CVML algorithms and in the “real” world, as people have been biased

[21]. Therefore, it is essential not only to reduce bias wherever possible but to also bring about a different understanding of how we trust technology and interpret its decisions.

Each of these harms seems like they can be solved technically by changing the algorithm to be “neutral.” However, these harms reflect underlying biases that exist in the world regardless of technology. Solving these biases will involve technical solutions, but it will also involve cultural and social solutions as well.

The reasons for bias are complicated. Sometimes it is as simple as the past being biased, and the algorithm is learning from past data. Other times, it has to do with dataset creation and neglect to include diversity. Sometimes, it is a technological issue related to different types of faces being harder to identify. And commonly, it is because researchers have false underlying assumptions about their results. However, all of the reasons behind bias boil down to one key concept: when programmers build these algorithms, they program in social values which cannot be neutral. Given that, it is necessary that we use an ethical framework to figure out the best way to reduce bias from the perspectives of each stakeholder.

Technologically speaking, there are several places that bias can be detected in a CVML system. Looking at the algorithm itself as somehow biased might be tempting. However, that is not the case. Algorithms are built from specific data, to solve specific tasks, and tested in specific ways, with specific values in mind. Algorithms are the only genuinely neutral part of the process; they are only doing what they have been told to do by programmers and the data they have been given. They have no moral agency.

Instead, the first real place to look is at the data that is training the algorithm. An example of this related to facial recognition is

a database that does not contain enough faces of people of color in order to be able to learn to recognize them. This is not as simple of a solution as it might seem, as many datasets that train algorithms begin to form worlds of their own that do not reflect the real world. A second but related place is the data used to test a machine learning system's accuracy. This could also be a set of data that lacks representation of all classes and so does not confirm that the algorithms work equally well for everyone.

Recently, researchers conducted a study on several datasets' images that are used to test algorithms created by universities [22]. The datasets were all created to represent similar objects, like databases of cars, or animals, and collected from the same source - the internet - but they all had drastically different reflections of the world. In fact, if one algorithm did well on one dataset - for instance, ImageNet - it would often experience a "dramatic drop in performance in all tasks and classes when testing on a different test set," like PASCAL VOC [22]. Adding more data to the datasets to test on did not improve the accuracy of the algorithm, but, instead, made them worse - indicating that the dataset itself was biased. The authors' conclusion at the end of this study is that "computer vision datasets are supposed to be a representation of the world. Yet, what we have been witnessing is that our datasets, instead of helping us train models that work in the real open world, have become closed worlds unto themselves" [22].

If this happens for small datasets that test algorithms, it also happens to large datasets that train algorithms that look at people. Even just thinking about the types of images that Facebook receives through its platform versus images uploaded to government websites versus images in the first search results from Google, it should be clear that those are all very different images of faces to train with - each not representing the world in its entirety.

A third way to isolate bias is in an algorithm result's interpretation. A controversial example of this is found in a Chinese study looking at criminality based on facial analysis of criminals and non-criminals called *Automated Inference on Criminality using Face Images* by Xiaolin Wu and Xi Zhang [23]. This article has received media attention, for good reason. Wu and Zhang attempt to distinguish criminals from non-criminals based



(a) Three samples in criminal ID photo set S_c .



(b) Three samples in non-criminal ID photo set S_n

Fig. 2. Sample identification photos from dataset [21].

on mere facial images. They gathered 2000 images of Chinese men between the ages of 18 and 55 with no distinguishing markings or facial hair and built four different supervised CVML classifiers (algorithms). See Fig. 2 for face samples supplied by the study. Unsurprisingly, the CNN (the neural network), did the best as it has been the source of many of the strides in CVML. Surprisingly, it was able to classify a Chinese man as a criminal with nearly 90% accuracy, just from his facial image.

Because of the controversial nature of this study, the researchers performed several validation techniques. They also re-ran the original experiment with random noise additions to the images to make sure that camera signatures were not causing any interference for determining criminality, which still produced statistically significant results [23]. They also tested the classifiers on random Chinese students with pictures they took themselves, and the results were again consistent.

The last part of the study was to determine which features of the face were consistently attributed to a criminal. They identified three structural measurements in the critical areas around eye corners, mouth, and philtrum that have significantly different distributions for the two populations, namely: the curvature of upper lip; the distance between two inner eye corners; and an angle between the nose and mouth [23]. In the end, they were able to isolate what the "average" criminal and non-criminal faces were in the database; they came up with three non-criminal faces and four criminal faces (see Fig. 3). Through a subjective test using real human judgment from 50 Chinese students, these seven faces appeared to agree with criminal/non-criminal human intuitions.

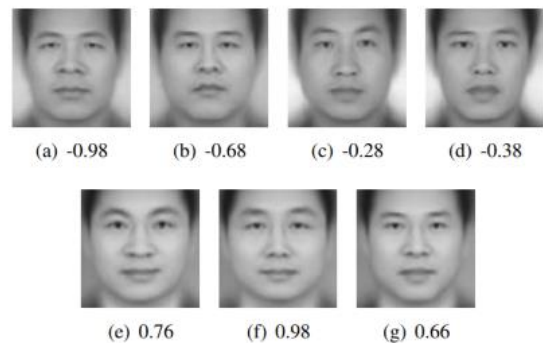


Fig. 3. (a), (b), (c), and (d) are the criminal average faces, (e), (f), and (g) are the three non-criminal average faces. The numbers below are representative of the score given from human judges (-1 for criminals, 1 for non-criminals) [21].

Did these Chinese researchers discover something important related to the nature of human physiology and its impact on criminality? Reading the research, it seems like they did a thorough job confirming their conclusions. Could it be true?

WIRED Magazine journalist Katherine Bailey wrote a response to this research study that fundamentally brings out the researcher bias that is involved in these Chinese researchers' interpretations [24]. She argues that the conclusions of the study are based entirely on the researchers' assumptions about how their society works. Bailey claims the study authors "simply assume there's no bias in the criminal justice system and thus that the criminals they have photos of are a representative sample of the criminals in the wider population (including those who have never been caught or convicted for their crimes)" [24]. This is an important point. If criminals were already stereotyped by their facial structure by the general population and more likely to be caught and convicted for

those reasons, then the photos they trained the algorithm with would already be biased in that way.

However, if “you start from the assumption that there isn’t any relationship between facial features and criminality... you are instead interested in whether there’s bias in the criminal justice system,” you would take the study results as evidence of such a bias [24]. You would not think that criminality is inherent in one’s facial features, but that people are biased against certain facial features in society. So, depending on a scientist’s fundamental assumptions and biases, the scientist may come up with an entirely different interpretation of the results.

Another point this article brings to the surface is the ramifications of any error rate if this research was actually applied to the criminal justice system in China. False positives would be catastrophic, not to mention the algorithm would likely have false negatives for many “non-criminal” faced people. This applies to the earlier discussion of error rates, but it shows how linked error rates and bias can be when it comes to real life applications of CVML systems, not to mention human rights.

This example shows that trying to create an “objective” approach to criminal profiling brings with it all the bias in a culture. If a society were actually to implement a system like this, it would structuralize the society’s bias and make it all the more difficult to change. That is why it is essential to implement CVML systems that are thoroughly vetted for harmful bias, if we implement them at all, in risky scenarios where false positives would be detrimental to society.

In each way that bias can be introduced, whether that be through training or testing data, or the interpretation of the data, the bias comes directly from the values of the creators of the algorithm. Whether or not that is a bias against a particular class of people, laziness, or fundamental assumptions about people, these biases can have a profound impact on an algorithm and its results.

Each of these issues is complicated, yet - they all have potential solutions when viewed through an ethical lens with each facet of society joining in to make each other accountable. The next section talks about the different parties involved in CVML, each responsible for making an ethical infrastructure for CVML to thrive.

IV. CVML STAKEHOLDERS

As already referenced, there are many stakeholders in CVML. Each of them has individual interests and responsibilities in the ethical infrastructure for CVML [26]. First, governments are interested in CVML primarily for issues of security, efficiency, and hegemony. Facial recognition technology promises to be a way to secure borders and protect against threats of terrorism as well as solving and preventing crimes on a local scale. It also is likely to help reduce inefficiencies in all areas of government by being a tool to verify identity. Additionally, facial recognition is a key aspect of the global AI “arms race” and promises to give significant benefits to the GDPs of nations that deploy it in various ways.

More than these substantial and quantifiable benefits, governments also have a vested interest in issues like Justice. They want to ensure fair use of facial recognition technology and that it is not targeted at people in a biased or harmful way. Governments have an obligation to encourage fairness and the Beneficent use of CVML.

Technology companies are another key stakeholder. As private and public companies, their primary motivator is profit. They want to build services that people will buy and use over their competitors. They have an interest in building long-term brands that are trustworthy and promote the common good, not public backlash. They also have a vested interest in remaining cutting edge and not falling behind their competitors, as well as maintaining proprietary algorithms. Technology companies also have a vested interest in promoting justice and reducing harms of their technology, even if it is just for profit. However, many companies do care about it for genuine moral reasons as well.

Society, in general, is often the consumer of AI, but it is also affected by AI in many ways. Whether that is in behind the scenes algorithms for a company’s hiring process or a person getting approved for a loan, they have an interest in algorithms being fair and reliable as well as not abusing their right to privacy and negatively affecting their identity. People in a democratic society can vote for elected officials and also have the ability to learn and investigate. More specifically in society, academic institutions have a responsibility in ethical research and implementations of AI as they are supposed to be unaffected by motives like profit. They have an interest in finding the best algorithms and discovering innovations in the name of research. They also have an interest in the education of AI.

Each of these stakeholders has specific responsibilities in the ethics framework of AI. They either work on creating ethical algorithms, work on holding people accountable who build them, or ethically use them. Often each stakeholder is involved in all three areas at some point. Sometimes it is easy to pass blame on others for failures in systems, but given the “overlap between social, political, commercial, and research interests” in AI, it would be a bad idea for “a single actor to have a monopoly on the ethics of AI and dominate the whole agenda” [25]. One stakeholder having control would create an environment where their interests are the ones put on the forefront, not everyone else’s. While, generally, everyone wants a better world, everyone wants that in a slightly different way and has a different part to play. If the United States government or a single company such as Microsoft becomes the only actor working towards ethics, their version of ethics is going to look very different than if all stakeholders have a voice in the discussion.

Taddeo and Floridi call this spread of responsibilities “distributed agency” as the “effects of decisions or actions based on AI are often the result of countless interactions among many actors” [1]. By giving everyone responsibility, and not just the government or tech companies, it “nudges all involved agents to adopt responsible behaviors” [1]. Floridi says that if we limit the “ethical discourse to individual agents” this “hinders the development of a satisfactory investigation of distributed morality” [26].

In order for each stakeholder to take responsibility and hold each other accountable, there has to be an established “ethical infrastructure” [26]. This means that the government, technology companies, and society must build systems that encourage trust, respect, reliability, privacy, transparency, freedom of expression, openness, and fair competition. Floridi argues that these ideas make the “morally good more likely to occur, and then become more stable and permanent, i.e., to take root” [26]. Building this ethics infrastructure is difficult, but it should be the goal of each stakeholder in CVML to prioritize the morally good and stabilize positive outcomes in the future.

Before going into specific actions that governments, technology companies, and society can take to build an ethical infrastructure and develop CVML with ethics on the forefront, it is essential to clarify how their individual responsibilities relate to one another. In a culture like the United States, it is sometimes unclear who is responsible for whose ethical behavior. If Floridi and Taddeo are correct, in every culture, everyone is responsible. However, that does not mean that each stakeholder must follow the same courses of actions or should be responsible for every aspect of CVML and ethics. Stakeholders still have specific jobs and parts to play in the process. If everyone were responsible for everyone else's actions, that would not make sense; however, if everyone is responsible for their attitude and a culture of openness and accountability and specific actions within that infrastructure, that makes more sense and is more practical.

A government's responsibility is building regulations and an environment that limits Maleficent behavior but encourages Beneficent behavior. They also must protect against attacks on Justice and Autonomy. This is a tall order and can get very complicated when people have different ideas of what these mean or they conflict. However, an example of this would be building laws that hold the correct people responsible for an AI mishap so that courts can enforce laws. As of now, there is no precedent for punishing a "learning machine" or recourse for an AI making wrong decisions. Government preemptively making decisions about this is wise.

Another example of government action is incentivization, whether that be giving positive incentives to follow specific standards or enforcing punishments for breaking others. They should work on incentivizing algorithms that are not easily twisted against civil rights, are accurate, and are bias-free. They also can incentivize algorithms that are Explicable – meaning that they are somewhat transparent and can be held accountable. This means encouraging research that helps people understand the ways that CVML makes decisions.

Governments also have a responsibility to be forward thinking and working on making sure that research in the area of CVML is moving ahead and that solutions to error rates and bias problems are being developed. This can look like a government research project or even grants to universities. This also includes bringing in experts to advise on future-proof regulations that will anticipate changes in the landscape to minimize harms.

Microsoft has agreed that technology companies have some responsibility towards ethics and has built many programs to address this, even if some criticize their lack of diversity. However, Microsoft believes that "it seems more sensible to ask an elected government to regulate companies than to ask unelected companies to regulate such a government" [7]. Technology companies should not have a responsibility to keep the government accountable; society and elections should do that. However, this does not mean that technology companies *cannot* hold government accountable, especially if they are incentivized by society to do so. Additionally, a government creating comprehensive regulation is "likely to be far more effective in meeting public goals" because "even if one or several tech companies alter their practices, problems will remain if others do not" [7]. Inherently, "the competitive dynamics between American tech companies – let alone between companies from different countries – will likely enable governments to keep purchasing and using new technology in ways the public may find unacceptable in the absence of a common regulatory framework"

[7]. Therefore, it is better for the governments of the world to lead the charge on regulating facial recognition for human rights and not technology companies; if a government or other entity pays a private company enough, they might be willing to create anything no matter what its potential harm.

However, while the government should lead the way on some aspects of ethically using facial recognition technology, it does not mean that technology companies are exempt from responsibility. They have an equal responsibility in building the ethical infrastructure. Brad Smith, for Microsoft, after arguing for government action, says that the "need for government leadership does not absolve technology companies of our own ethical responsibilities... [tech companies] have a responsibility to ensure that this technology is human-centered and developed in a manner consistent with broadly held societal values" [7]. In the United States, tech companies are the main stakeholders working towards an ethical infrastructure at the moment, as the current federal administration has done little to regulate or guide CVML. Tech companies like Microsoft have been the leading lobbyists for facial recognition laws and the entities leading the charge to protect civil liberties, while the government has mostly been taking a backseat. The administration is mostly asking how they can get out of the way of innovation, not asking how they can work to protect citizens from abuse.

Brad Smith has published several blog posts and co-authored a book that recommends courses of action for government entities and technology companies when dealing with facial recognition technology specific to protecting human rights [7] [27]. In the book *The Future Computed*, co-written by Smith and which tackles more than just facial recognition, a critical section is on privacy and security within AI. Among other things, this section suggests that the industry itself must develop standards to comply with values of privacy and keep track of how consumer data is used in different steps of AI training and deployment, even if it is not government mandated [27].

Technology companies, beyond advocating for ethics around AI in government practices, should be working on best practices and standards for CVML. They also should be working to make algorithms as understandable as possible to the public without severely limiting their competitiveness. The largest area of ethics tech companies directly control is the development and deployment of algorithms. Building infrastructures in their own companies that promote accurate and unbiased algorithms is a big part of this responsibility.

Government and technology companies have significant responsibilities, but this does not let consumers and other institutions off the hook. Society, in general, has responsibilities as well, especially when it concerns the accountability of the government and technology companies in building infrastructures. In the United States, consumers have a significant influence on technology companies themselves because of market forces and consumer backlash when things go wrong. Consumers also have a deciding factor in government decisions as they elect representatives who will hopefully care about these issues. Consumers, NGOs, and even academia need to be involved at the ground floor to set norms and expectations about how they want AI to interact with people and human rights.

Society also needs to demand transparency. Transparency is not a silver bullet, but it does allow for society to gauge who and what needs to be held accountable, and it also allows society to become a part of the conversation. Governments and technology

companies need to work towards this, but society absolutely needs transparency to interact within this distributed morality framework. Without knowledge about what is going on, society is often helpless to make decisions and act out their Autonomy.

Society's other responsibility is to bring in voices that are not represented in technology companies and the government – voices that are usually the most affected by biased and error-prone algorithms. Minorities and people in the fringes need to be in the conversation in order for harms to be discovered and for solutions to be built.

Overall, the stakeholders mentioned above have serious responsibilities. However, there are some stakeholders, like society, that have a difficult time engaging if there are not actions taken by the government and technology companies. Additionally, there are some functions that only government can do – like pass and enforce laws.

V. SOLUTIONS AND ETHICS

Each issue that is specific to CVML, namely, human rights concerns, error rates, and bias, can be addressed most effectively in a distributed responsibility framework through the lens of bioethics. The next part of the discussion will go through different possible solutions and how they measure up in an ethical framework. These solutions are not supposed to be comprehensive. Instead, they should give guidance on how to think about applying the bioethics framework when considering a policy or action.

A. Responsibilities to Protect Human Rights

The common reasoning behind court decisions regarding privacy and technology in the United States is a person's "expectation of privacy" [28]. For government searches, the Fourth Amendment is used to determine "reasonable search." Future court decisions for new technologies will use these two premises as their basis. However, determining what someone should consider their expectation of privacy or what constitutes reasonable search in the age of AI is difficult. The courts have no way of being able to measure evolving "expectations" and mostly rely on their own opinions of what is reasonable to determine this. There is court precedent from past technologies, but none have the same potentials as AI and big data. None deal directly with a person's identity in the same way. Courts making arbitrary determinations can be dangerous and could mean that a gap in privacy protection exists for users of CVML until serious abuses are uncovered and backlash ensues. If a court case with serious implications for consumer protection has to make its way into the court system up to the Supreme Court to offer protections, this leaves a lot of people vulnerable for a very long time.

Making ethically based laws from the beginning is a better solution than reacting to potential government or corporation abuse of power via the Supreme Court or the backlash of public opinion. This makes the job of courts simpler as they must interpret current laws and not rely on the Fourth Amendment or User Agreements. These laws would have to be followed from the beginning and would eliminate confusion and abuses from the start. While many states and municipalities in the United States are developing AI laws, San Francisco and New York are good examples to bring into the discussion because of their different

approaches to predict the uses of AI and protect against its potential abuse.

San Francisco recently passed a city-wide ordinance to ban government organizations from using facial recognition technology in its entirety [29]. The ordinance also forces government agencies to get permission before placing any new surveillance technologies by a new review board. This is an extreme measure to reduce government abuse of CVML. However, the law does not address consumer-facing companies and their use of facial recognition. So, if a break-in were to happen at a grocery store, the government could not use facial recognition technology on the surveillance cameras, but the grocery store could [29]. Arguably, this may be considered a form of backlash to government surveillance in the past, but it also is a backlash against the harms that other forms of AI have already caused in the news. Facial recognition is entirely prevented from being used, even if it could help solve or prevent crimes, because of fear of the technology's misuse.

In applying the bioethics principles, this law seems to be addressing concerns of Maleficence that may come up. The city is concerned that the technology could be used for nefarious reasons and for invading citizen privacy. Additionally, with the new processes and permissions required for any new type of surveillance technology, the city is also targeting Explicability by requiring both accountability and transparency. However, this ordinance is eliminating any Beneficence that could happen with facial recognition. San Francisco has decided that facial recognition, with its inherent error rates and bias issues, is too high of a risk to give to law enforcement. Instead of trying to come up with mitigating techniques for the risks, like making sure a facial recognition match could not be enough to convict someone of a crime, they have decided to ban it entirely.

New York City has proposed a bill to regulate biometrics on consumers by businesses, requiring businesses to post warnings and URLs to further information if they are collecting biometric data such as face images [30]. This addresses issues of consent with regards to facial recognition. Consent is important to protect human rights because it allows customers to gauge their own privacy risk and weigh it against the services or products of a business. Consent is essential to Autonomy. This law would also make it so companies would have to be honest and upfront with what they would be doing with the data and how they would secure it.

This law in New York is more preventative than reactive. While specific businesses and complaints against them may have triggered it, it is not banning the practices they use – it is just regulating them to interact with public interests. It seems in this instance that New York is working towards coming up with solutions that balance government interests, business interests, and consumer interests. This bill also cares about preventing Maleficent behavior, but it also is allowing for Beneficial behavior as well. The bill is using the means of Explicability and Autonomy (user consent) to try to push the use of these technologies to positive outcomes and not covert and negative ones. If this bill is passed, executed correctly, and followed through is made in enforcement and investigations, this law could prevent civil rights abuses from businesses and allow customers to be aware of any risks from the beginning. This hopefully will prevent any backlash that could ensue from a business using facial recognition data without a customer's knowledge.

While some individual states and cities are creating regulations in the United States, federal engagement is necessary. A company like Google already has trouble fitting its technology to each country's individual privacy regulations; having different state or city regulations within that will be a huge hassle – especially if part of their technology is banned altogether in some regions. The federal government should look at ways to consistently regulate CVML to eliminate the need for states to regulate it themselves in haphazard ways. Setting some basic standards and protections on a federal level, both for government agencies and private companies, would go a long way in preventing abuse and giving citizens both choices and protection.

Worldwide, each society is served best by protecting civil liberties on the onset and not waiting for pushback from the public. The alternative will inevitably lead to civil rights violations and possibly an overcorrection when regulations are applied in the future. Without some baseline for civil rights, the race to the most utilized facial recognition algorithm with the best deployments will be a race to the bottom. Instead, all governments should provide “a floor of responsibility that supports healthy market competition. And a solid floor requires that we ensure that this technology, and the organizations that develop and use it, are governed by the rule of law” [8].

Using the Explicability principle should be a primary guiding factor for government intervention, as it will allow for the principles of Autonomy to be used by consumers and technology companies. Working on reducing Maleficent behavior while advocating for Beneficent behavior is a delicate balance, but it can only be done in Explicable conditions where there can be conversations that bring in multiple stakeholders. This should ultimately lead to Justice where the balance maintained is fair and non-biased for all and protects everyone's civil rights.

All that has been addressed so far are legislative approaches to protecting human rights from facial recognition abuse, and early ones at that. These are necessary for the successful integration of facial recognition into society, but they are not adequate. Technology companies cannot have the attitude towards civil liberties of merely following the letter of the law. The law cannot predict every possible harm and prevent them from happening. Instead, tech companies and society need to start from an ethical framework that instills values related to civil liberties and protecting people from harm.

One recent example of a technology company deciding on facial recognition is Microsoft refusing to supply facial recognition technology to California police departments for their police body cameras and dash cams [31]. California, as a state, wanted to run a scan every time an officer pulled someone over. Brad Smith said Microsoft rejected the opportunity because of the human rights concerns related to error rates and bias. Microsoft is working hard to reduce these but does not have confidence that they are reduced enough for California to use facial recognition with negligible risk. Brad Smith also mentioned they refused to sign a contract with an unnamed country who wanted to use their facial recognition technology to spy on people in their capital city [31]. This example shows that a technology company can make an ethical choice based on their own risk assessment, but it also shows that just because one company refuses to participate, does not mean that another will not. It is possible that another company, like Amazon, Google, or Face++, may decide that they want to contract with the State of California.

Another goal that technology companies can work towards is giving more people voices in the discussion. While several technology companies and organizations have built AI ethics boards to help advise development and deployment of AI technologies, there has been a problem with representation within the panels. It is good that they are trying to get feedback, but all too often voices representing communities that would be affected the most from error rates and bias are missing [32]. Microsoft, who has in many ways been leading the ethics discussions around facial recognition in the industry and who has requested government regulation, not only suffers from a lack of diversity within their own company but also has created a research group without any African American members [32]. An ethics-focused industry group called the Partnership on AI, launched by Google, Amazon, IBM, and Microsoft does not have any African American board members or staff listed online and has a board predominantly made up of men [32]. Academic institutions also suffer from this; Stanford recently announced a new artificial intelligence institute with specific goals for it to represent all of humanity. However, the original 120 members of the institute did not include a single African American [32].

While it is a known fact that white men dominate technology companies in Seattle and Silicon Valley, it is unfortunate that this is carrying over into ethics conversations because those who are the most marginalized and who possibly would be able to foresee risks and bias in future algorithms are excluded from these conversations. A recent report by the AI Now Institute found that only 15% of AI researchers at Facebook and 10% of AI researchers at Google are women [33]. And, in general, only 2.5% of Google's workforce is black, while Facebook and Microsoft are at 4% [33]. This report also mentioned several recommendations on what technology companies could do to improve diversity [33].

A lack of diversity is problematic from a distributed responsibility framework, as stakeholders who should be given responsibility - and a say in decisions - are not given it. This is also problematic from an Explicability standpoint, as the accountability component cannot be carried out without problems being pointed out and addressed by those who are most affected. This also undermines the Justice component because fairness is not a high standard in the process.

In bioethics, this is easier to accomplish as society and patients have more extensive access to what they are interacting with because it deals directly with their own body. The added ethical component of Explicability that AI4People advocates as a key fifth ethical principle is vital to society being able to interact with AI decisions [4]. Society needs to understand how and why CVML is used to be able to hold tech companies and governments accountable. This is the fundamental mechanism for society to gain access to the discussion table. However, as of now, society is very out of the loop. Academics and some NGOs are working on educating people about risks of AI, but they do this without access to actual tech company algorithms and government algorithms for the most part. They are looking and speculating without real access to how things are working. Governments may be able to help in requiring more ability for users to consent and understand how AI is affecting their lives, like in the case of New York's proposed law. However, technology companies are going to have to be willing to cooperate.

Microsoft argues that it is in their best interest to work with the public on any new facial recognition technology they deploy, but

how much access to the inner workings of their systems does this grant the public? There is a balance between maintaining patent law and keeping proprietary algorithms competitive, while still being able to audit them to make sure that they are protecting the best interests of society. This balance will have to be explored.

B. Responsibilities to Reduce Error Rates

Just as Explicability is vital to protecting civil rights, it is also essential for society to understand AI systems in order to interpret what error rates mean and how they impact the limitations of a system. This understanding or lack of understanding can affect Justice as well. Microsoft thinks that "[a]n AI system could also be unfair if people do not understand the limitations of the system, especially if they assume technical systems are more accurate and precise than people, and therefore more authoritative" [27]. This brings up a key point about how people view technology versus how they view a person doing the same action. When a person identifies another person, society believes them, but understands human limitations for memory and identification. Additionally, humans can also lie. However, when facial recognition is used, humans often trust it unconditionally. This may be because of crime television shows infallibly using this technology. However, this is wrong. Facial recognition has error rates too, even if it is getting to be more reliable than people. If people knew this, they might treat the results differently and look to other avenues to corroborate the truth.

This could be even worse than a system result being wrong: blindly trusting CVML can hand our responsibility over to machines and remove some of our moral agency. These "technologies can inhibit our moral agency when we abdicate our responsibilities by unreflectively outsourcing our authority to digital assistants and algorithms" [3]. When algorithms are making important decisions and society does not even consider their moral weight, that is a recipe for disaster.

One idea for government intervention to promote Explicability is creating national standards for facial recognition that have bench marks for error rates and bias. Rep. Emanuel Cleaver (D-Mo.) submitted a letter to NIST asking for them to create standards, especially as facial recognition relates to demographic differences in error rates [34]. He also asked NIST to investigate data sets used by facial recognition developers and come out with demographic standards for representation. This could be an opportunity to reduce bias as well as error rates. In terms of NIST standards, giving a government certification for a facial recognition technology may be a way to gain public trust in the technology. Even if this test is still voluntary, a company may be required to undergo the certification to compete for a government contract. This action would both let consumers in on the error rates that are inherent in facial recognition, while at the same time, it would give a clear indication of what algorithms were accurate and which were not. The tricky part is figuring out what real "accuracy" means and what datasets are representative of the real world in a way to accurately test that result. For instance, if the error rates of a benchmark test are low, they still may not translate into the real world – which could give false confidence.

Technology companies can be a part of this process and work to create standards that they feel are possible to achieve and that can be easily tested. Additionally, if technology companies decide to "opt-in" to testing, they show that they are committed to accuracy.

At the same time, society should be encouraging technology companies to be a part of this standardization process while working on building standards that reflect their interests. Academia and NGOs need to be involved in the process; it should not just be a conversation between technology companies and the government.

Another action that may need to be made is deciding if there are some domains that false positives would be too risky to make. For many uses of facial recognition, like tagging photos on Facebook or even unlocking a smartphone, a false positive is relatively low risk for the person identified. However, risk increases the more decisions are made based on this information and how "certain" it is believed this information is. If people can be convicted of a crime on just a facial recognition match from a CCTV, that seems like a very risky false positive. If someone can be flagged in error as dangerous or a criminal in a police investigation or interaction, that places potential undue harm on the flagged person.

The public needs to be able to evaluate these use cases and decide whether the error rate and consequences for false positives make these applications too risky. In San Francisco, as mentioned earlier, they have made that decision. While it may be a bit overkill in some people's minds, it does protect from false positives. However, it does so at the cost of potential uses for facial recognition that would help solve and prevent crimes with low risk for false positives. For instance, in solving crimes after the fact, it might be helpful to use facial recognition to see if there are any likely suspects based on an image. It might not be allowed as a way for the police to arrest someone or courts to convict someone, but it could give them a way to begin investigating. This seems like a balance that considers multiples stakeholders' interests, produces potentially positive outcomes, and reduces the potential for Maleficence. It also allows law enforcement to maintain Autonomy by both having tools that help them investigate at their disposal but also being able to choose and keep moral responsibility while using them. Hopefully, these factors will lead to an increase in Justice.

C. Responsibilities to Eliminate Bias

As this paper has discussed, no matter what, values are going to be represented in an algorithm. We have seen that "[t]here is danger in thinking of technology as simply neutral. Human agency is involved in the design and use of all technologies: a designer's intentions shape a technology, and its efficacy is complicated by a user's intentions" [3]. If we continue to let the default be society's underlying values, "the default tendency of these systems will be to reflect our darkest biases" [18]. However, there is no way to "neutralize" an algorithm of its creator's values. However, maybe the values of creators may be able to be shaped to more closely resemble Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability.

Can we neutralize the training data or limit interpretations of results? This is tempting, but impossible [18]. If we consider bias as a purely technical problem, we are already missing part of the picture. Bias is a social issue first and a technical one second [18]. Governments, technology companies, and society are going to have to work on fixing social problems while they work on building technical solutions to bias, which will also be required.

An example of a possible way this could work is a new law that the New York City Council just passed [35]. Their AI accountability bill places transparency requirements on

algorithms used by the government [35]. This bill has a specific focus on bias and figuring out which algorithms affect marginalized communities in unfair ways, including those used for school placement and police resource distribution. This bill is not banning or requiring anything for algorithms specifically, but it is taking a step forward in investigating what laws should be in place to protect citizens from harm. The bill also builds a task force that includes representatives from the departments who use algorithms, members of technical industries, and technical ethicists. This bill is forcing the New York City government, tech companies, and society to work together and for each stakeholder to have their voice heard.

This task force is also tasked with figuring out how to alert residents to when they are subject to an algorithm's reach, like when an algorithm makes decisions about where to dispatch police officers in different parts of the city. Additionally, the task force is also looking at data that trains the algorithms to see if there is a way to make it more public and to analyze it for bias.

This bill takes a lot of positive actions from the perspective of bioethics. It brings different stakeholders together while promoting various means of Explicability. While the government is leading this action, it still requires the cooperation of different parts of the technology industry and the academic side of society. It also allows for Beneficent government programs to stay in place, while looking out for Maleficent outcomes and outcomes that are Unjust for certain parts of the city. It also addresses Autonomy and Explicability with the public by giving them warning about when different algorithms are affecting them.

Some other avenues for instilling ethical principles to prevent bias can come from technology companies. Like mentioned previously, diversity is a key component of accountability and Justice on ethics advisory councils. It is also essential for diversity to be on the teams that make CVML – having someone on the team from a minority background increase the perspectives on the algorithm and help it be built to avoid representational harms. It also prevents interpreting results with biased assumptions, as people with different perspectives on life often have different assumptions as well. Technology companies can create environments with a diversity of perspective by hiring diverse teams on purpose for these types of projects.

Another avenue technology companies can improve their chance at reducing bias in algorithms is by encouraging third-party testing and auditing before deployment of the technology. Technology companies should not be afraid of bias found at this stage; they should be afraid of it appearing after deployment and millions of people are using it. By opting into NIST tests and other tests that exist, as well as welcoming academics into test algorithms themselves, it will save technology companies from pain and backlash later.

V. CONCLUSION

As is evident, it is difficult to completely pull apart the stakeholders and assign them specific tasks and responsibilities. While it is easy to say that the government needs to lead the charge in some instances and to regulate facial recognition consistently, it is hard to say how they actually can accomplish this without the help of tech companies and society. Tech companies have to be willing to cooperate and be on the same page in terms of wanting to protect consumer privacy and other civil rights. Society has to be willing to elect people that will make these types of decisions

that will serve their best interests and to pressure tech companies to be more transparent and invite consumers, NGOs, and/or academics to be a part of the process of development and deployment. This might seem like a crazy, far-fetched utopia of cooperation, but it has happened in small pockets of the tech industry already and can continue to happen if each party recognizes their own part to play in the process.

CVML has impressive potential to save lives, like in New Delhi, where authorities were able to use new facial recognition technologies to find 3,000 missing children in just four days [36]. Facial recognition has also been used to diagnose rare genetic diseases that have facial markers [37]. Additionally, facial recognition has the potential to completely change how we handle and secure money, as card-less ATMs are in development with card-less shopping centers already in testing around the world [38]. In order to promote these types of applications and make them representative of the future, ethics needs to be built into facial recognition from the beginning. This will only be the norm if governments, corporations, and society work together to build an ethical infrastructure that promotes Explicability and strives towards Justice.

ACKNOWLEDGMENT

I would like to thank Professor Elaine Weltz and Dr. Carlos Arias from the Seattle Pacific University Computer Science Department for their input into this project and guidance throughout my time at SPU. They have each helped me cultivate my passion for this subject matter as well as provided invaluable feedback. I would also like to thank my family and friends for their encouragement throughout the researching and writing process.

REFERENCES

- [1] M. Taddeo and L. Floridi, "How AI Can Be a Force for Good," *Science*, vol. 361, no. 6404, pp. 751-752, 2018.
- [2] B. A. Hamilton, "HOW DO MACHINES LEARN?," Booz Allen Hamilton, [Online]. Available: <https://www.boozallen.com/s/insight/blog/how-do-machines-learn.html>. [Accessed 24 February 2019].
- [3] M. J. Paulus, B. D. Baker and M. D. Langford, "The Digital Transformation of Higher Education: A Framework for Digital Wisdom," *Christian Scholar's Review*, vol. XLIX, no. 1, Fall 2019.
- [4] L. Floridi, J. Cowsls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke and E. Vayena, "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689-707, 26 November 2018.
- [5] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, 9 October 018. [Online]. Available: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>. [Accessed 26 2 2019].
- [6] Imperial College London, "Written Submission to House of Lords Select Committee on Artificial Intelligence [AIC0214]," 2017.
- [7] B. Smith, "Facial recognition technology: The need for public regulation and corporate responsibility," 2018.
- [8] B. Smith, "Facial recognition: It's time for action," 2018.
- [9] J. Collins, "China is using facial recognition to track millions of Muslim citizens wherever they go," Quartz, 17 February 2019. [Online]. Available: <https://qz.com/1552708/china-is-using-facial-recognition-to-track-millions-of-muslim-citizens-wherever-they-go/>.
- [10] P. J. Grother, M. L. Ngan and K. K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification," U.S. Department of Commerce, 2018.
- [11] J. Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots," American Civil Liberties Union, 26 July 2018.

- [Online]. Available: <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>. [Accessed 22 March 2019].
- [12] S. Liao, "Chinese facial recognition system mistakes a face on a bus for a jaywalker," *The Verge*, 22 November 2018. [Online]. Available: <https://www.theverge.com/2018/11/22/18107885/china-facial-recognition-mistaken-jaywalker>. [Accessed 22 3 2019].
- [13] All Tech Asia, "Face++: the company that's looking at your face behind Meitu and Alipay," 16 February 2016. [Online]. Available: <https://medium.com/act-news/face-the-company-that-s-looking-at-your-face-behind-meitu-and-alipay-ab30c24716b7>.
- [14] B. B. Watch, "Face Off: The lawless growth of facial recognition in UK policing," May 2018. [Online]. Available: <https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf>. [Accessed 20 May 2019].
- [15] J. C. Fox, "Brown University student mistakenly identified as Sri Lanka bombing suspect," *The Boston Globe*, 28 April 2019. [Online]. Available: <https://www.bostonglobe.com/metro/2019/04/28/brown-student-mistaken-identified-sri-lanka-bombings-suspect/0hP2YwyYi4qrCEdxKZCpZM/story.html>. [Accessed 20 May 2019].
- [16] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Conference on Fairness, Accountability and Transparency*, pp. 77-91, 2018.
- [17] T. Simonite, "When it Comes to Gorillas, Google Photos Remains Blind," *Wired*, 2018.
- [18] K. Crawford, "The Trouble with Bias," *Neural Information Processing Systems*, Long Beach, 2017.
- [19] T. Bolukbasi, K. Chang, J. Zou, V. Saligrama and A. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing world embeddings," *Advances in neural information processing systems*, pp. 4349-4357, 2016.
- [20] O. Reference, "Exnomination," Oxford University Press, Oxford.
- [21] J. Lanier, *You are not a gadget: A manifesto*, New York: Knopf, 2010.
- [22] A. Torralba and A. A. Efros, "Unbiased look at Dataset Bias," *Computer Vision and Pattern Recognition*, pp. 1521-1528, 2011.
- [23] X. Wu and X. Zhang, "Automated inference on criminality using face images," *arXiv preprint arXiv:1611.04135*, pp. 4038-4052, 2016.
- [24] K. Bailey, "Put Away Your Machine Learning Hammer, Criminality is Not a Nail," *Wired*, 29 November 2016. [Online]. Available: <https://www.wired.com/2016/11/put-away-your-machine-learning-hammer-criminality-is-not-a-nail/>. [Accessed 27 March 2019].
- [25] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo and L. Floridi, "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach," *Science and engineering ethics*, vol. 24, no. 2, pp. 505-528, 1 April 2018.
- [26] L. Floridi, "Disturbed Morality in an Information Society," *Science and Engineering Ethics*, vol. 19, no. 3, pp. 727-743, 2013.
- [27] Microsoft, *The Future Computed: Artificial Intelligence and its role in society*, Redmond: Microsoft Corporation, 2018.
- [28] N. LaChance, "At Supreme Court, Debate Over Phone Privacy Has A Long History," *NPR*, 16 March 2016. [Online]. Available: <https://www.npr.org/sections/alltechconsidered/2016/02/29/468609371/at-supreme-court-debate-over-phone-privacy-has-a-long-history>.
- [29] S. Fussell, "San Francisco Wants to Ban Government Face Recognition," *The Atlantic*, 5 February 2019. [Online]. Available: <https://www.theatlantic.com/technology/archive/2019/02/san-francisco-proposes-ban-government-face-recognition/581923/>.
- [30] S. J. Kapinos, "New York City Considers Facial Recognition Bill — Will New York Be the Next Forum for Biometric Privacy Litigation?," *Sr National Law Review*, 31 January 2019. [Online]. Available: <https://www.natlawreview.com/article/new-york-city-considers-facial-recognition-bill-will-new-york-be-next-forum>.
- [31] M. Moon, "Microsoft didn't want to sell its facial recognition tech to California police," *engadget*, 17 April 2019. [Online]. Available: <https://www.engadget.com/2019/04/17/microsoft-facial-recognition-california-police/>. [Accessed 23 April 2019].
- [32] S. Levin, "'Bias deep inside the code': the problem with AI 'ethics' in Silicon Valley," *The Guardian*, 29 March 2019. [Online]. Available: <https://www.theguardian.com/technology/2019/mar/28/big-tech-ai-ethics-boards-prejudice>. [Accessed 30 March 2019].
- [33] S. M. West, M. Whittaker and K. Crawford, "Discriminating Systems: Gender Race, and Power in AI," April 2019. [Online]. Available: <https://ainowinstitute.org/discriminating-systems.pdf>.
- [34] E. C. II, "Congressman Emmanuel Cleaver, II," 26 November 2018. [Online]. Available: https://cleaver.house.gov/sites/cleaver.house.gov/files/NIST_LETTER_FR_T.pdf. [Accessed 23 March 2019].
- [35] J. Vacca, H. K. Rosenthal, C. D. Johnson, J. R. Salamanca, V. J. Gentile, J. R. E. Cornegy, J. D. Williams, B. Kallos and C. Menchaca, *Automated decision systems used by agencies.*, New York: The New York City Council, 2018.
- [36] P. T. o. India, "Delhi: Facial recognition system helps trace 3,000 missing children in 4 days," *The Times of India*, 22 April 2018. [Online]. Available: <https://timesofindia.indiatimes.com/city/delhi/delhi-facial-recognition-system-helps-trace-3000-missing-children-in-4-days/articleshow/63870129.cms>. [Accessed 14 January 2019].
- [37] J. Mjoseph, "Facial recognition software helps diagnose rare genetic disease," *National Human Genome Research Institute*, 2017.
- [38] Microsoft, "NAB and Microsoft leverage AI technology to build card-less ATM concept," Microsoft, 2018.



Abagayle L. Blank Student at Seattle Pacific University studying computer science, information studies, and communication. She will be employed at Microsoft as a Software Engineer post-graduation. Professional interests include Artificial Intelligence, ethics, computer science, information theory, and theology.

APPENDIX

Governments, technology companies, and society each have parts to play in creating a future where CVML and facial recognition are beneficial for all. As a Christian in computer science, is there an even more specific responsibility for me? After looking at bias, it is evident that the systems that I will create are going to reflect my beliefs about the world and I will be training my own biases into any programs I build. Knowing what I believe, why I believe it, and understanding how that interacts with society is immensely important as I also develop systems that impact people.

Technology can be used to create many things. I could design a website for a non-profit or help make an app to aid in reducing homelessness. There is also a dark side to computer science; using people for monetary gain, invading privacy, locking people out of progress, and misrepresenting people. There are definite choices that need to be made from an ethics standpoint, and often, a technology invented for one purpose can be exploited for a negative one. Understanding these choices and knowing that things are not often black and white is essential to the scholarship in my field.

My honors project is an example of this as I tackle how a technology that aids society in convenience, security, and other uses can further racism and bias if it is used without care. This project aims to show that technology can be used for good, but also evil. It is the job of a Christian scholar in computer science to aid in targeting technology for its wise uses.

This connection to scholarship is echoed by George Marsden when he says that Christian scholarship involves that we “do what we can to promote the cause of the light and to use all our talents where they may be helpful” [1]. We cannot be arrogant about human knowledge and technology, but our “Christian belief should be a source of humility” [1]. I agree with this sentiment; Christian scholarship requires that we know that our talents can be used for a specific purpose and that our human knowledge needs to be used in humility lest we make mistakes out of arrogance.

Another viewpoint on scholarship that I have investigated comes from *Scholarship and Christian Faith: Enlarging the Conversation*. This helped me to position myself in a scholarship tradition, after not being sure how to label myself coming from a non-denominational background. What I found was that there is a non-denominational tradition of scholarship that I could see myself in as a teenager [2]. This tradition “centers on the Bible alone and on the need always to start afresh... Ideas are not to be handed down from the past but rather to be discovered anew” [2]. However, I do not see myself in this place anymore. I am more and more relying on theologians from the past to inform my faith, and I depend every day on other people’s discoveries in computer science to aid my journey there as well. I do not understand the need always to reinvent the wheel, and this becomes painfully obvious in programming as well.

Instead, I see myself in more of the Wesleyan tradition. Using the quadrilateral of the Bible, tradition, experience, and reason to inform my scholarship instead of relying on my interpretation of the Bible or the facts around me [2]. I see myself “situated in larger contexts of relationship and conversation” [2]. This perspective has allowed me to appreciate my part of the conversation but also understand that I am part of a much larger picture and can learn from everyone around me.

In turn, I also see myself relying on others in computer science to inform how I view making the right decisions as it relates to CVML. I am not trying to reinvent how to look at ethics but have adapted many other people’s ideas into my project and applied them to CVML specifically.

The last aspect of scholarship that I would like to discuss is that of application. My field is very much one that is driven by use and implementation. Beyond that, my honors project is looking at ethics from a consequentialist viewpoint, not focusing on people’s “good intentions.” Finding an intersection with this and my faith has been more challenging for me. God only cares about what is in my heart, right?

It turns out, many people have talked about the relationship between intentions, actions, and effects in Christianity. One of the principles that has resulted is Tomas Aquinas’ doctrine of *double effect* [3]. The basic premise is that intentions and consequences both matter for an act to be considered morally good, and that the benefits can outweigh any evil that may come about because of an action. This can be related to the ethical principles in my honors project, as the principles of beneficence, non-maleficence, and justice mean that the good must outweigh the bad and the good must be as equally distributed as possible. However, my honors project does not quite go as far as to say that the intentions have to be good for all parties. For instance, not *all* technology companies have to have good intentions if the government is regulating their actions correctly for good to occur in society. A good AI society has to be thinking that there will be people with bad intentions.

However, I do need to worry about having good intentions as a Christian scholar and matching those with wise and good actions. And, for the most part, stakeholders with good intentions are those that have the best actions. Therefore, intentions are important, even if it is consequences that are focused on. So, while it is great to start with good intentions, and somewhat necessary, real and right solutions need to be the result.

Overall, Christians have a responsibility in the ethics of AI. They are not just a part of society, which has specific responsibilities, but they also have certain expectations placed upon them by their holy calling. As a Christian, I should be using my gifts in a way that benefits people, be working with past voices in my discipline to inform my decisions, and think about my intentions as well as my actions’ potential outcomes.

REFERENCES

- [1] G. Marsden, *Outrageous Idea of Christian Scholarship*. Cary: Oxford University Press, 1997, pp. 83-100.
- [2] D. Jacobsen, R. Jacobsen and R. Sawatsky, *Scholarship and Christian faith*. Oxford: Oxford University Press, 2004.
- [3] A. McIntyre, "Doctrine of Double Effect", *Plato.stanford.edu*, 2004. [Online]. Available: <https://plato.stanford.edu/entries/double-effect/>. [Accessed: 24- Apr- 2019].