Artificial Intelligence Art: How artificial intelligence can affect the current bias within facial recognition systems

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Jaehoon Smith

Spring, 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Bryn E. Seabrook, Department of Engineering and Society

Artificial Intelligence Art: How artificial intelligence can affect the current bias within facial recognition systems

Introduction

Artificial Intelligence (AI) art has opened an infinite outlet of creative direction for the art industry. From creating innovative architectural designs to curating stimulating food presentations for the culinary world, AI art has the ability to revolutionize creative thought processes at the click of a button (AIWS, 2020). However, AI art maintains the ability to serve in other productive industries such as crime regulation specifically through facial recognition. Facial recognition systems are not simply an innately just tool as improper use of this technology can lead to many societal issues. The Wicked Problem Framing supports this notion that facial recognition systems and AI art cannot serve as the immediate solution to all facial detection problems (Buchanan, 2019).

Research Question and Methods

Given the recent technological growth within the past decade, many believe AI or facial recognition to simply exist prominently in science fiction movies. However, both AI and facial recognition systems are quite prevalent in today's world and are utilized in multiple industries such as healthcare, law enforcement, or social media. Given the recent rise of Artificial Intelligence art, how could it affect the current bias within facial recognition systems? The key aspect that binds these technological systems is the notion that they are created by humans and subsequently, human bias. The algorithms that drive these technologies require diverse perspectives and databases to reduce biases within society. While still new, AI has shown precedence in its societal impacts, both good or bad, through many case studies. The potential

power of AI must be analyzed and policies must be cautiously made to improve its growth rather than becoming a technology to be feared.

Background Information

Facial recognition systems work by using deep learning algorithms to analyze visual data and identify individuals. The system compares the visual data to a database of known faces to find a match and improve accuracy by focusing on key facial structure data points such as the distance between the eyes and the shape of the nose and mouth. As a result, the system outputs data such as age, gender, and facial expressions to produce a match. AI art programs, such as DALL-E, can receive a visual or textual semantic, generate images similar to the prompt, and produce a completely original image based on that prompt. These programs have the ability to alter existing images and narrow down millions of images into one (O'Connor, 2023).

The utilization of AI art programs can help facial recognition databases to better analyze incoming data and optimize their algorithms to improve accuracy. Machine learning is a critical component of facial recognition technology, as algorithms are programmed to "learn" from previous experiences to solve problems that are too complex to solve with conventional programming (Pxl, 2023). The creation of AI art with variations of people in every race, age, or gender will provide images for facial recognition systems to practice and improve their identification process. By gaining experience in matching features of artificial images, facial recognition systems can learn to better identify a target match even if their appearances have changed.

In addition to improving accuracy, the use of AI art also addresses the issue of privacy. Even though an existing photo of the match target must exist, artificial photos of "faces that do

not exist" allow for databases to not intrude on people's privacy. Facial recognition technology maps and analyzes face geometry and facial expressions to identify a similar match. The technology typically scans the distance between eyes, nose, mouth, etc. and adds to its memory of how a certain archetype of people may look (Amazon, 2020). AI art provides a safe and vast database that facial recognition systems can use to minimize privacy concerns. By utilizing AI art, facial recognition systems can improve accuracy while also addressing privacy concerns, making it a crucial component in the development of these systems.

STS Framework: Wicked Problem Framing

The integration of AI art and facial recognition systems addressing the Wicked Problem Framing could have both positive and negative impacts. Wicked problems are complex and difficult to solve, and traditional analytical approaches are not effective in addressing them. Additionally, wicked problems have no stopping point for a solution and must continuously grow alongside societal discussions. “The search for scientific bases for confronting problems of social policy is bound to fail, because of the nature of these problems” (Rittel & Webber, 1973). These problems can lead to unintended consequences, whereas tame problems have well-defined solutions with clear stopping points that can be evaluated as right or wrong.

As a positive, AI and facial recognition technology could provide more efficient and accurate methods for identifying individuals and facilitate the implementation of solutions to varying problems, such as public safety and crime prevention. The deep learning method of facial recognition could also be used to analyze large amounts of data, such as demographics and social behaviors, to better understand the root causes of these inaccuracies and to identify patterns that could lead to more effective solutions. However, the usage of AI art and facial

recognition systems in addressing these wicked problems also raises concerns about privacy and civil liberties. The use of these technologies could result in widespread surveillance and the collection of personal data without consent, leading to a loss of privacy and freedom. Additionally, facial recognition systems have been criticized for their lack of accuracy, particularly with regard to race, gender, and age.

Furthermore, the deployment of AI in general raises ethical and capability questions about the role of the technology in solving complex social issues. Similar to many systems, every AI contains the paradox of artificial agency which is described as the "paradoxical effect that artificial agents, learning autonomously from human-derived data, will often learn human biases—both good and bad" (Osaba & Welser IV, 2017, p17). This paradox articulates the idea that AI may be a constant growth in knowledge yet derives its information strictly from the source of human intelligence, no matter the quality. This key aspect factors into the inherent bias of humans, which contains racial, political, and religious beliefs that alter each input of data. In policing systems, AI has already served as prominent cases for intaking misinformation. In 2019, Amara Majeed was misidentified as a Sri Lanken Bomber to the public and police forces later announced their mistake (Fox, 2019). Majeed had not been recognized as a fully accurate match, yet due to human confidence, the AI system took the information as a valid match and future data point.

This issue of inherent bias is further worsened by the widespread adoption of similar predictive policing in many cities across the United States. As further implementations of AI are constantly influenced by its bias derived from human input, the question arises whether the use of AI is a wicked problem. While the experimental use of AI proves effective, there is no clear

answer whether AI will become a stable solution but rather as a complex technology that balances between good and bad.

Results and Discussion

The current state of Artificial Intelligence has reached an all time high point in today's technological society and must be categorized as a Wicked Problem Framing. As AI grows in interest within the technological focus of modern times, the pace at which AI is growing is alarming compared to the policies that garner the proper usage of it. The process in which AI slowly becomes a wider aspect of everyday life is approaching to become a wicked problem. Currently, AI is regarded as imperfect and inhuman as no matter how well programmed an AI responds to humans, it is unlikely to develop actual emotional attachment that provides the proper context within any situation (IQMotion, 2022). As such, AI is a technology of constant growth through reinforced data and the information it is provided draws from the compilation of human bias. Given its current time and diverse information, AI has already become an everpresent feature of everyday life and therefore must be monitored to improve as a diverse system. In fact, AI currently serves in many aspects of societal industries such as healthcare, law enforcement, and social media.

Healthcare AI

Priced at a hefty $808 billion industry in the United States, healthcare is an essential part of American life (Zauderer, 2023). With such vital decisions for patients occurring constantly, healthcare centers utilize AI to parse through data in mimicking medical decisions of doctors. However, AI fails in its algorithms in detecting blind spots of various races of human beings due

to its inherent human bias. Researchers and physicians who use the algorithms have determined a significant racial bias in determining kidney transplants, heart surgeries and breast cancer diagnoses that affect the health of millions of patients (Fleur, 2022). In 2019, a clinical study found many hospitals employing an algorithm that deemed black patients had to be deemed much sicker than white patients in order to be recommended for the same care. This issue arose from the algorithm referencing its data from past healthcare spending trends. The trend indicated that black patients typically spent less on healthcare than white patients due to the history of wealth and income disparities (Obermeyer, 2019). Another algorithm allocated the number of hours an Arkansas resident with disability should receive care for. In-home care residents had their hours cut to an extreme and was later revealed to be an error in the medical priority levels for people's location, regardless of disability (Lecher, 2018).

Like many flaws in AI systems, the hospital AI algorithms lack the database of information that allows for proper determinations. Based on outdated information generated over decades and human error of negligent coding, AI systems develop their conclusive results with the given input of people. AI's dependence on its inherent coding is not a good or bad trait. The implementation of AI in healthcare is designed to speed up processes such that more people can receive care. Healthcare AI is not meant to include the biases to exclude marginalized patient groups. The issue of healthcare AI extends from the issue of data inconsistency. The use of AI and algorithmic decision-making systems are ran with little policing from any higher organization. Human lives are at risk with every decision of a machine, yet much of the algorithms and their exact processes can be unknown to even the clinicians (Grant, 2022). The effectiveness and usability of the algorithm is not called into question but rather the bias that influences final decisions.

Organizations such as the U.S. Department of Health and Human Services (HHS) and the Food and Drug Administration (FDA) are in charge of regulating medical devices, including medical AI systems. These tools are vital for patient care as they detect diagnosis/treatment and reduce time spent in clinical and administrative issues. However, these technologies that provide decisions in risk of mortality, likelihood of readmission and off ground treatment, are not required to be reviewed or consistently regulated by any organization.

Developers of electronic health record (EHR) software, Epic Systems, released an AI tool for detecting early signs of Sepsis and became a common implementation in over 170 health centers. The tool findings revealed that it detected 7% of patients correctly for early signs of sepsis but did not detect 67% of patients who later developed sepsis (Richardson, 2021). Epic Systems later criticized the findings, indicating that the sepsis model was not correctly calibrated to the specific patient population and backed their AI with multiple studies proving its performance. The issue then arises in the concept that hospitals need to tailor the system to their specific environment. Attempting to implement this tool would call for retraining of staff, calibrating the AI with specific demographic data, and many more logistical concerns. If a system works in California, but not in New York, the key difference of performance lies in the lack of the AI understanding the diverse populations within each state. Simply transferring AI from one context to another without strict supervision of population or systemic differences can introduce inherent significant bias (Price II, 2019). The abundance of these AI systems and their impact are accentuated when development is encouraged but not regulated under a proper organization. The unclear transparency and the heavy adjustment of AI healthcare systems lacks the regulation to sustain the usage of AI making decisions on the level of an experienced clinician.

Law Enforcement Facial Recognition

Facial recognition is currently employed in many ways for both commercial and law enforcement organizations. On the commercial side, companies such as IBM, Amazon, and Microsoft have developed their respective facial recognition technologies that have been since discontinued due to their inaccurate results among categories such as race, age, and gender. A 2018 research found that lighter skinned participants had an inaccuracy of 0.8% while black women had 34.7% (Buolamwini et al, 2018). This significant difference among recognition was tied to the lack of diverse data sets that represented the marginalized groups.

On the opposite spectrum, law enforcement bases their facial recognition algorithms on mugshot databases, which is predominately more prevalent with images of black people. Specifically within the U.S., black people have higher arrest rates than white people for similar crimes (Balko, 2022). This indicates that law enforcement facial recognition systems draw their data points from another database that lacks diversity for its readings. Facial recognition systems are not a perfected form of technology, but the root of the misidentification problem is the lack of non-marginalized and diverse databases.

The process in which law enforcement facial recognition operates has another layer of inherent bias. In the case of the U.S., this bias stems from historical racism pertaining to racial tensions risen from surveillance of marginalized groups. The 18th century lantern law demonstrated the systemic targeting which is as follows, "...laws in New York City that demanded that Black, mixed-race and Indigenous enslaved people carry candle lanterns with them if they walked about the city after sunset, and not in the company of a white person. The law prescribed various punishments for those that didn't carry this supervisory device."

(Garcia-Rojas, 2016). In comparison to modern times, police surveillance disproportionately monitor marginalized communities.

Facial recognition algorithms are effective to a fault. Both commercial and law enforcement applications of facial recognition maintain the ability to detect facial features based on their databases yet the differentiating factor is the implicit bias of humans inputting fractionary representation of data points. It is meant to capture the good aspect of law enforcement yet lacks the bandwidth of diversity to be effective. With so many different groups of culture, race, and age, AI must address many independent factors that make it seem impossible to solve its inherent bias. Whether a person only develops relationships within their own race, avoids certain demographic areas, or labels an acquaintance based on their inherent traits, implicit human bias has an effect on many life decisions (Grinberg, 2015). When such bias is impacting an algorithm that only provides results based on the information it is given, potentially life changing systems like facial recognition need much attention and regulation.

Social Media

As a much more relatable topic, social media has introduced the use of algorithms to alter the content that is displayed to its users. From notifications of trending media to targeted displays of advertisements, social media culminates data based on your entire lifestyle by using an algorithm. However, as much as the algorithm displays information, it also has the ability to censor and limit content consumption. Instagram received criticism for being biased against accounts that aligned with philosophies such as LGBTQ+, political, etc.

The act of filtering this data was coined the term "shadowbanning" and social media platforms were able to control a user's social media content in such a way that they are unaware that it is happening (Brown, 2021). While the algorithm is designed to engage the user with characteristic content designed solely for the user, the fundamental aspect of AI is that it learns and changes based on the algorithm of its developers. By the influence of implicit biases of the developers, the algorithm that is unable to understand human behavior and emotion will extract this key aspect from the code that the developers implemented. Many of these social media platforms were initially influenced by the algorithm implemented by a dominant white male culture. Twitter, for example, has the history of prioritizing right-leaning content or displaying notable demographic differences in preview images (Wilton, 2021). Social media users are unable to ascertain as to what information is available to them and the algorithms display content without a proper analysis of the user's identity, preferences, and beliefs. Given so many aspects of a user's likeness, there is no definitive formula to finding the perfect algorithm solution to include and accommodate every user.

Fundamentals of Facial Recognition and AI Art

When it comes to scanning a human face, most facial recognition systems search for the following facial features: distance between the eyes, forehead to the chin, the nose and mouth, depth of the eye sockets, shape of the cheekbones, and contour of the lips, ears, and chin (Maheshkar et al, 2012). Yet even with these numerous measurements, these features do not cover much of the possible human face. Much is to be considered, whether there is makeup used, usage of accessories or obstructing facial hair and color. Given the extent of possible categorical features, facial recognition systems must prioritize narrowing key facial data points to determine

the most defining traits of age, race, and gender. Compiling these facial characteristics can provide diverse data points considering the potential bias of current facial recognition systems. Skin tone, eye shape, and expression are unique to different faces and have high levels of variance. Accessories provide another level of variance as items such as earrings or sunglasses can hinder facial features and reduce the accuracies of other data points (Min et al, 2011). The vast variations of a human face requires facial recognition databases to become more diverse and continuously develop reinforced heuristics to increase facial recognition accuracy.

AI art provides a potential method of populating facial recognition databases. AI art images have a much broader range in specifying traits, leading to a wider range of potential images. AI images can specify certain traits of a face which increases the deep learning process of the AI art algorithm that constantly builds upon itself (Sahar Mor, 2022). To depict the potential of AI art, Fig. 4 represents a ”portrait of an 18 year old white female wearing lemon sunglasses posing in front of a white car wearing a green shirt and black jacket”.

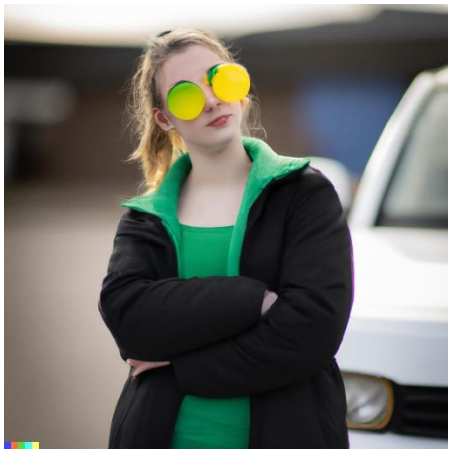

Fig. 4. portrait of an 18 year old white female wearing lemon sunglasses posing in front of a white car wearing a green shirt and black jacket

Despite the long description and grammatically confusing request, AI generation has the ability to fill the gap of missing information within facial recognition databases. What currently

lacks for most facial recognition systems are the available images to improve the reinforcement learning system on. AI art bridges this issue by offering an infinite database of images for the program to practice, increase overall accuracy, and recognize common facial patterns.

Although AI art has the potential to reduce societal effects of bias and misidentification, it still pertains to the discussion of human bias and its developmental state requiring proper regulation. Just like any AI system, AI art requires human prompt in order to generate a result. AI is neither a good or bad system as it only analyzes the data that it is given. Implicit bias and stylistic wording of a human alters the prompt that is inputted into the generation of a potential face for a facial recognition database. As a result, even utilizing an AI model to aid another AI model still holds the inherent flaw of human bias. Furthermore, the lack of policy on AI art raises yet another issue. If a facial generation is identical to an existing person or malicious content is pipelined into the potential facial recognition database, there is no current regulation to prevent such issues.

Future of AI

Without a doubt, AI has yet to be a fined tuned tool for consistent and professional work. AI, however, maintains the budding advantage of infinite variations of data. Much is left to be refined in data collection and respective policies must be established to reduce algorithmic impact. With the extreme growth of AI, it is vital to keep track of its development in all fields while ensuring it does not remain unregulated. Although, as it stands now, AI is incapable of taking decisive stances and factoring emotion into it's responses. Whether it be due to the limitations of technology or the ethical issues of applying emotion to inanimate entities, AI has no ability to act at the full contextual capacity of a human. AI is a compiler of textual evidence

but still holds the integral component of human input. The issue of AI's codependency with humans exemplifies the bias that is rooted within AI analysis. Many forms of AI now lack a diverse database due to the information it is provided by humans and therefore must require a diverse set of human input in order to accommodate for all potential bias. Further research could explore the combination of the various forms of AI. While AI can be applied in many fields, the key aspect shared across all industries of these algorithms currently focuses on human traits. As more AI develops, more study can be done in reinforcing these databases to function as a whole.

Conclusion

AI art has proven the ability to impact facial recognition, although there still remains much to improve and regulate before becoming an integral method of diversifying databases. AI is a relatively young industry with a promising future for reducing human workload and simplifying lifestyle processes. As AI slowly becomes integrated into human lives, it is important to appoint regulations that match the fast pace at which AI is growing. AI maintains the position of becoming a vital technology for society yet must address critical issues before it is able to operate along with the intricate nuances of human nature.

Works Cited

AIWS (2020, November 11). *Some unbelievable facts about AI in arts*. AIWS. . Retrieved February 5, 2023, from https://aiworldschool.com/research/some-unbelievable-facts-about-ai-in-arts/

Amazon Web Services. (n.d.). *What is Facial Recognition?*. Amazon. Retrieved February 5, 2023, from https://aws.amazon.com/what-is/facial-recognition/#:~:text=It%20works%20by%20identifying%20and,large%20collection%20of%20existing%20images.

Balko, R. (2020, June 10). *Opinion | there's overwhelming evidence that the criminal justice system is racist. here's the proof.* The Washington Post. Retrieved March 22, 2023, from https://www.washingtonpost.com/graphics/2020/opinions/systemic-racism-police-evidence-criminal-justice-system/#Misdemeanors

Brown, A. (2022, November 9). *Understanding the technical and societal relationship between Shadowbanning and algorithmic bias*. Forbes. Retrieved March 22, 2023, from https://www.forbes.com/sites/anniebrown/2021/10/27/understanding-the-technical-and-societal-relationship-between-shadowbanning-and-algorithmic-bias/?sh=6499f5766296

Buchanan, R. (1992). *Wicked Problems in Design Thinking.* MIT Press. https://web.mit.edu/jrankin/www/engin_as_lib_art/Design_thinking.pdf

Buolamwini, J., & Gebru, T. (2018, January 21). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. PMLR. Retrieved March 22, 2023, from http://proceedings.mlr.press/v81/buolamwini18a.html

Fleur, N. S. (2022, August 29). *'racism is America's oldest algorithm': How bias creeps into health care AI*. STAT. Retrieved March 22, 2023, from https://www.statnews.com/2022/05/30/how-bias-creeps-into-health-care-ai/#:~:text=Researchers%20and%20physicians%20have%20warned,health%20of%20millions%20of%20patients.

Fox, J. C. (2019, April 28). *Brown University student mistakenly identified as Sri Lanka bombing suspect*. BostonGlobe.com. https://www.bostonglobe.com/metro/2019/04/28/brown-student-mistaken-identified-sri-lanka-bombings-suspect/0hP2YwyYi4qrCEdxKZCpZM/story.html

Garcia-Rojas, C. (2016, March 3). *The surveillance of blackness: From the trans-atlantic slave trade to Contemporary Surveillance Technologies*. Truthout. Retrieved March 22, 2023, from https://truthout.org/articles/the-surveillance-of-blackness-from-the-slave-trade-to-the-police/

Grant, C. (2023, February 24). *Algorithms are making decisions about health care, which may only worsen medical racism: ACLU*. American Civil Liberties Union. Retrieved March 22, 2023, from https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism

Grinberg, E. (2015, November 25). *4 ways you might be displaying hidden bias in everyday life*. CNN. Retrieved March 22, 2023, from https://www.cnn.com/2015/11/24/living/implicit-bias-tests-feat/index.html

IQMotion. (2023, February 17). *Can human be replace by Artificial Intelligence?* IQ MOTION. Retrieved March 22, 2023, from https://www.iqmotion.ai/human-vs-ai/#:~:text=No%20matter%20how%20well%2Dprogrammed,is%20vital%20to%20business%20growth.

Lecher, C. (2018, March 21). *What happens when an algorithm cuts your health care*. The Verge. Retrieved March 22, 2023, from https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Maheshkar, V., Agarwal, S., Srivastava, V. K., & Maheshkar, S. (2012). *Face Recognition using Geometric Measurements, Directional Edges and Directional Multiresolution Information* . ScienceDirect. Retrieved March 22, 2023, from https://www.sciencedirect.com/science/article/pii/S2212017312006597?ref=pdf_download&fr=RR-2&rr=7abdaedf7e6c82c8

Min, R., Hadid, A., & Dugelay, J.-L. (2011, April). *Improving the recognition of faces occluded by facial accessories*. ResearchGate. Retrieved March 22, 2023, from https://www.researchgate.net/publication/224238159_Improving_the_recognition_of_faces_occluded_by_facial_accessories

Obermeyer, Z. (2019, October 25). *Dissecting racial bias in an algorithm used to manage the ... - science*. Science.org. Retrieved March 22, 2023, from https://www.science.org/doi/10.1126/science.aax2342

O'Connor, R. (2023, January 7). *How dall-e 2 actually works*. News, Tutorials, AI Research. Retrieved February 5, 2023, from https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Osaba, O. A., & Welser, W. I. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR1744.html

Price II, W. N. (2019, April 1). *Medical AI and Contextual Bias*. SSRN. Retrieved March 22, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3347890

Pxl. (2023, January 12). *Machine learning and face recognition: PXL Vision*. Machine learning and face recognition | PXL Vision. Retrieved February 5, 2023, from https://www.pxl-vision.com/en/blog/machine-learning-and-how-it-applies-to-facial-recognition-technology

Richardson, L. (2021, December 16). *Artificial intelligence can improve health care-but not without human oversight*. The Pew Charitable Trusts. Retrieved March 22, 2023, from https://www.pewtrusts.org/en/research-and-analysis/articles/2021/12/16/artificial-intelligence-can-improve-health-care-but-not-without-human-oversight

Rittel & Webber. (1973). *Dillemas in a General Theory of Planning*. Retrieved February 6,

2023, fromhttps://www.sympoetic.net/Managing_Complexity/complexity_files/1973%20Rittel%20and%20Webber%20Wicked%20Problems.pdf

Sahar Mor, S. (2022, April 16). *How dall-E 2 could solve major computer vision challenges*. VentureBeat. Retrieved March 22, 2023, from https://venturebeat.com/datadecisionmakers/how-dall-e-2-could-solve-major-computer-vision-challenges/

Wilton, T. (2021, October 22). *Twitter says its platform promotes right-leaning content.* ScreenRant. Retrieved March 22, 2023, from https://screenrant.com/twitter-platform-promotes-right-leaning-content/

Zauderer, S. (2023, January 11). *39 U.S. Healthcare Industry Statistics, facts & size*. Life-Changing ABA Therapy - Cross River Therapy. Retrieved March 22, 2023, from https://www.crossrivertherapy.com/healthcare-industry-statistics#:~:text=The%20healthcare%20industry%20is%20worth,revenue%20comes%20from%20patient%20care.