Bias in Facial Recognition: Causes and Effects in Policing

Luca Manolache

Palo Alto High School

AP Seminar

Ms. Filppu

Bias in Facial Recognition: Causes and Effects in Policing

Introduction

Artificial intelligence (AI) is being used in increasingly many places to aid in decisions (Livingston, 2020). Specifically, one significant area adopting AI is policing. However, we argue that this increase will only further existing biases and should not be used in its current state. While there are numerous definitions of AI, in this paper, AI will be referred to as deep neural networks (DNN). Similarly, facial recognition refers specifically to an implementation of facial recognition involving a DNN. A DNN at a basic level is an algorithm that attempts to find the best mathematical function to predict from given data (Pei et. al., 2017). Internally, in a DNN, thousands to millions of mathematical functions are applied to the input to reach a specific output. To find these functions, a neural network requires a large amount of labeled data. In facial recognition, labeled data is an image of a face with a description which the DNN will try to learn. In most cases, to create a DNN, one simply needs to procure labeled data. However, it is incredibly difficult to look into how the DNN internally makes the decisions it does, which is needed for finding biases. When a DNN is used for a facial recognition system, it can be used for many tasks by police including "individual and crowd interaction systems used [for] surveillance… [which] are useful for spotting sociopathic or destructive types of behavior (e.g., car theft, arson, mugging, robbery, vandalism)" (Nunn, 2001). However, while these uses are beneficial to police and public safety, research shows that AIs can contain "learned biases towards physical properties like skin tone and facial complexion" (Fuchs, 2018). As a point of clarification, we will define bias as "a tendency to believe that some people, ideas, etc., are better than others that usually results in treating some people unfairly" (Merriam-Webster, n.d.).

Causes of Bias

Most academic sources are in consensus that facial recognition has biases. The primary causes are bias in training data, insufficient amounts of training data, lack of regulation, and the very math behind the DNN algorithms. As Roselli and Matthews, part of the department of Computer Science at Clarkson University explains, since AI learns from datasets that have been labeled by humans or from historical data, "[u]nfortunately, this includes learning human biases contained therein" (Roselli et. al., 2019). If an AI is fed data that contains biased or false information, it will learn that information and use it in its future predictions. This is an obvious concern in policing, a system that is often criticized for its bias, as past and present human biases can be taught to the facial recognition systems. A similar issue that arises is that more data exists for individuals that are part of the majority. Since an AI learns to predict based on the data it is given, if the data is from mainly one race, it will work far better on people of that race. Roselli notes this, "commercial facial recognition systems trained on mostly fair-skinned subjects have vastly different accuracies for different populations: 0.8% [error] for lighter-skinned men and 34.7% [error] for darker-skinned women" (Roselli et. al., 2019). This is not always the intentional fault of the developers and companies creating these systems as the data can be difficult to assemble and check for contained biases. Still, as police are known for having racial bias (West, 2018), this could serve to only increase that. Another potential cause of bias that often compounds with the previous point is when data is missing. Missing data creates unlearned cases; this is specifically a problem when the data is missing a certain group of people, as even models with high accuracy will not be able to accurately identify unlearned cases (Roselli et. al., 2019). This is particularly damaging to minorities as they are often not well represented by the data. As such, this can have devastating consequences, for instance, Bradford, part of the Institute of Security and Crime Science at University College London, gives an example:

"[i]magine …an algorithm that selects nursing candidates for a multi-specialty practice—but it only selects white females" (Bradford et. al., 2020). In policing, this could manifest as an AI that is more likely to misidentify minority ethnic groups, leading to a further increase in bias. It has been noted by researchers that "facial recognition algorithms often disproportionately misidentify minority ethnic groups and women" (Bradford et. al., 2020). Given that "Physical technologies like body-worn video (BWV), drones, GPS and enhanced scanning equipment are being used more and more" (Bradford, 2020), the inevitable increase in facial recognition is worrying, as it will cause minorities to be misidentified more by police and increase bias.

Data-driven biases are compounded by a lack of regulation upon companies selling machine learning algorithms. McClellan, states that the private sector has no regulations surrounding facial recognition and the government has few, unclear regulations (McCellan, 2019). The lack of regulation makes finding this bias more difficult as there is little reason or motivation for companies to spend large amounts of time and money attempting to locate the stated biases. Even with regulation, the difficulty of finding biases is a large, unsolved issue due to the difficulty in understanding the inner workings of facial recognition algorithms (Xie et. al., 2019). Because DNNs are represented as a mathematical function with thousands to millions of different parameters, when looking at the internals, it is incredibly difficult or impossible to understand what each parameter does and how it impacts the decisions the system makes. DNNs are, therefore, given the term "black box," a system in which you feed input and get an output without knowing what happens inside of the system" (Pei et. al., 2017). The difficulty in understanding DNNs makes creating regulations near useless as proving that one system works correctly can be near impossible. If police agencies use such technologies, they have no way of knowing that the product is unbiased. One area which could be tested to infer if a DNN is biased

is the training sets (Roselli et. al., 2019). However, this unfortunately does not provide a completely accurate understanding of how the AI makes decisions or prove it is unbiased (Xie et. al., 2019) and companies do not always make their training sets public. For policing, this means that the agencies or the public can not scan the training sets themselves and even if they could, this would not be sufficient to prove a lack of bias.

Solutions

Despite the seemingly bleak outlook, solutions may exist to improve this issue. Creating or giving an existing government institution the power to regulate AIs would be a step in the right direction, however, as mentioned above, this is not always possible or effective. One possible regulation could be giving the police or another government institution, such as the National Institute of Standards and Technology, the authority to look into private companies' training data (without releasing it to the public) to scan for potential bias. While not a perfect fix, police could employ this alongside other proposed tools such as simulating faces to create new test cases (McDuff, 2018) or using algorithms to provide an easy summary of a DNNs internals (Xie et. al., 2019). However, as these tools are rather new and not fully researched, they could give potentially meaningless outputs. Therefore, it is essential that the police not rely on facial recognition software due to its bias until a well-researched solution can accurately verify the lack of bias.

Works Cited

Bradford, B., Yesberg, J. A., Jackson, J., & Dawson, P. (2020). Live facial recognition: Trust and legitimacy as predictors of public support for police use of new technology. The British Journal of Criminology. https://doi.org/10.1093/bjc/azaa032

Fuchs, D. J. (2018). The dangers of human-like bias in machine-learning algorithms. Missouri S&T's Peer to Peer, 2(1), 1.

Livingston, M. (2020). Preventing racial bias in federal AI. Impacts of Emerging Technologies on Inequality and Sustainability, 16(02). https://doi.org/10.38126/JSPG160205

Merriam-Webster. (n.d.). Bias definition & meaning. Merriam-Webster. Retrieved February 22, 2022, from https://www.merriam-webster.com/dictionary/bias

McClellan, E. (2019). Facial Recognition Technology: Balancing the Benefits and Concerns. J. Bus. & Tech. L., 15, 363.

McDuff, D., Cheng, R., & Kapoor, A. (2018). Identifying bias in ai using simulation. arXiv preprint arXiv:1810.00471.

Nunn, S. (2001). Police technology in cities: changes and challenges. Technology in society, 23(1), 11-27.

Pei, K., Cao, Y., Yang, J., & Jana, S. (2017). DeepXplore. ACM Symposium on Operating Systems Principles. https://doi.org/10.1145/3132747.3132785

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. Proceedings of the 2019 World Wide Web Conference. https://doi.org/10.1145/3308560.3317590

West, J. (2018). *Racial Bias in Police Investigations*. Retrieved February 22, 2022.

Xie, X., Ma, L., Juefei-xu, F., Xue, M., Chen, H., Liu, Y., Zhao, J., Li, B., Yin, J., & See, S. (2019). DeepHunter: A coverage-guided fuzz testing framework for deep neural networks.

ACM SIGSOFT International Symposium on Software Testing and Analysis, 28.

https://doi.org/10.1145/3293882.3330579