THE KENAN INSTITUTE FOR ETHICS at DUKE UNIVERSITY

Duke PRATT SCHOOL of ENGINEERING

Case Study

FACIAL RECOGNITION SOFTWARE

Matthew Mosca, Amber Díaz Pearson, Stacy Tantum

This case study was developed with support from the Lane Family Ethics in Technology Program and the Kenan Institute for Ethics. It was completed under the direction of Dr. Amber Díaz Pearson, The Kenan Institute for Ethics, and Dr. Stacy Tantum, Department of Electrical & Computer Engineering, Pratt School of Engineering, Duke University.

Please direct any questions to Stacy Tantum at stacy.tantum@duke.edu

FACIAL RECOGNITION SOFTWARE

The following is based on fact, but organizations and software mentioned are fictional.

In 2015, a large tech company, Fluvian, released affordably-priced commercial facial recognition software. Their software is one of several cutting-edge facial recognition systems on the market today.

The system architects for Fluvian's facial recognition project did not design the software toward any particular application. They instead hoped that potential clients would see the software as a powerful tool that could be applied in many different areas. The goal was to build an affordable and powerful software tool that would be cost-effective, even for smaller organizations with smaller budgets.

The data scientists, machine learning engineers, and software developers at Fluvian who were responsible for designing and programming the facial recognition system needed a database of facial images for training the algorithms. With a deadline fast approaching and a tight budget, the project leaders opted for the cheapest and most easily attainable database. Some data scientists noted a lack of diversity among the faces in the database – the people in the database were overwhelmingly male and of lighter skin tones. Once the machine learning engineers were made aware of this potential issue, some of them expressed concern about how the lack of females and individuals with darker skin tones in the training database could impact the software's performance, but they also knew that assembling or obtaining a database that better represents the diversity of our population would require time and money they did not have. The software developers assured the rest of the team that after the completion of the first iteration of the facial recognition system they could improve the system using a better training database, so the team pushed on. As a compromise, the team decided to make available to end users the match-confidence percentage, a quantitative measure of how well the target photo resembles the matched photo from the database, for each photo match returned by the system.

Since Fluvian's facial recognition software's public release, it has been the subject of significant criticism claiming algorithmic bias. A 2018 study found that the software was more accurate for lighter-skinned and male faces, and less accurate for darker-skinned and female faces, which suggests it exhibits gender and racial bias. The study pointed to a lack of diversity in the training database as a likely contributor to the apparent bias, because other companies that developed similar software had previously cited non-diverse training databases as a problem they had faced. However, since Fluvian considers their algorithm to be proprietary, it does not share information regarding the details of its software's algorithms and development, including the training database.

Due to the lack of transparency from Fluvian, the study could not be certain of the root causes of the apparent bias in the facial recognition software. Fluvian's CTO has denied the relevance of the 2018 study to the accuracy of its facial recognition software, citing problems with the study's methods.

One of the early adopters of Fluvian's facial recognition software was a police department in Anytown, USA. A recent news article reported that the police department has been using Fluvian's software to identify and apprehend suspects by comparing a photo, or artist's sketch, of the suspected offender to a photo database of everyone who has been arrested in the city. A programmer who worked in the police department built an internal user interface to present a simple display of the facial recognition software's results for the officers. While Fluvian's software provided a match-confidence percentage upon completion of its search, the programmer decided not to show this number and opted to instead always display the top five matches, regardless of the level of confidence in each match. As a result, five results are always shown, even if the match-confidence percentage is low, indicating the match does not resemble the photo, or artist's sketch, that was uploaded. The only rules officers must follow (when using the internal system built on top of Fluvian's facial recognition software) were written by the police department.

Critics have argued that since the police department's user interface always returns the top five matches, even if the resemblance to the photo in the database is poor, it changes the question from "Who committed this crime?" to "Did this individual commit this crime?" and this contributes to biased outcomes. Furthermore, they argue that it makes people who have been arrested in the city previously more likely to be arrested for future crimes than those who have never been arrested in the city. Also, due to the study showing algorithmic bias in Fluvian's software, there is a concern that women and people of color are at risk of being disproportionately misidentified and suspected of crimes they did not commit.

Researchers have argued for federal regulation of facial recognition technology, to ensure that it is not misused. While some cities have entirely banned the use of facial recognition technology by the government for any purpose, including law enforcement, no national regulations currently exist. Fluvian executives continue to emphasize positive applications of the technology and argue that it should not be banned or significantly hindered in its development simply because there may be a possibility that it could be used inappropriately.

Questions for reflection:

1. What factors are most important in your evaluation of the current status and implementation of facial recognition software?
    a. Is the technical accuracy and reliability of the software important?
    b. Do the motivations of the company's leadership matter?
    c. Do openness and algorithmic transparency matter?
2. Do you agree with any of the critics? If you agree with any of the critics, which ones do you agree with and why? If you do not agree with any of the critics, why not?
3. What is the good produced by current implementations of facial recognition software? What is the potential good from future implementations of this type of software?
4. What harms are produced by current implementations of facial recognition software? What are other potential harms from future implementations of this type of software?
    a. What steps could be taken to mitigate existing harms and prevent future harms? How would you decide which steps, if any, to take?
5. What are the intended and unintended consequences of current implementations of facial recognition software?
    a. Do current implementations contribute to a common good, or do they disproportionately distribute benefits to some and burdens to others?
6. To what parties did the developers at Fluvian have obligations when they designed the facial recognition software and made decisions about the training database?
7. To what parties did the police department programmer have obligations when they designed the internal user interface and made decisions about how to display the matches to the photos in the database?
8. What do you think about the character of the different persons in the case (including software developers, data scientists, engineers, and the police department programmer)?
    a. Do you think anyone acted honorably? With concern for others? Out of selfish motives?