Think Data

Principles: Fairness and non-discrimination, Privacy, Human oversight and determination, Responsibility and accountability, Awareness and literacy, Multi-stakeholder and adaptive governance and collaboration
Values: Human Rights , Ensuring diversity and inclusiveness, Living in harmony and peace
Stakeholders: Civil society, Public sector, Media, Private sector, Academia, Technical community

Accounting for racial abuse on smartphone's facial recognition system

Google updates its photo gallery application by adding a facial recognition software to classify similar pictures together under a tag. The goal is to help the user in finding pictures form its gallery quicker. The software is included in all the latest updates by Google worldwide.

The software uses Google's data as reference to compare the pictures and classify them accordingly when it finds similarities between the picture and the reference. However, no thorough data biases check was done.

When a young African American man uploads a selfie that he had taken, he is shocked to find that his pictures had been classified and tagged in the "Gorillas" category by the system. The error goes viral and similar cases appear where "non-white" people are wrongly tagged and compared to animals.

Google decided to delete the Gorilla section from its tags. The data that had been given as a reference was full of biases that were seen as "normal" by the software. Because most of the data that has been gathered comes from predominantly white western countries, image recognition software struggle in correctly analyzing darker faces as well lighter skin tones.

The ethical implications of facial recognition are widespread, as evidenced in this scenario. The facial recognition technology used in this scenario enhanced racial discrimination of a part of the population as it had not been adapted to the demographic it could monitor.

How could this situation have been avoided in the first place?

Here are our recommendations per stakeholder:

- Programmers have to work on longer-term improvements in the use of words to label photos and its image recognition software, which automatically generates the tags. The samples used to gather data must be diverse enough to provide an accurate representation of reality.

- Human resources are also encouraged to hire more diverse people so that different

perspectives can be added.

- The civil society and the media are encouraged to share stories of such type so that Google and other companies can better their data.

- Government and International / Regional Organization should encourage more people into working in the field of AI, such as woman and people of different origins or physical attributes as to reduce the data biases. To do so, education programs for these target populations are developed to raise awareness and interest on the topic.

Privacy & Human Rights & Responsibility and accountability & Safety and security & Fairness and non-discrimination

Know more about this case:

- "Google rushes to fix software that tagged photo with racial slur",
  CNN, https://edition.cnn.com/2015/07/02/tech/google-image-recognition-gorillas-tag/index.html

- Google apologises for Photos app's racist blunder,
  BBC, https://www.bbc.com/news/technology-33347866

- Google's solution to accidental algorithmic racism: ban gorillas", The
  Guardian, https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people

Related work:

- "Why algorithms can be racist and sexist",
  Vox, https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency

- "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms",
  Brookings, https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best

-practices-and-policies-to-reduce-consumer-harms/

- "Facial Recognition Is Accurate, if You're a White Guy", New York Times, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html