IBM didn’t inform people when it used their Flickr photos for facial recognition training

IBM took nearly a million photos from Flickr, used them to figure out how to train facial recognition training programs, and shared them with outside researchers. But as NBC points out, the people photographed on Flickr didn’t consent to having their photos used to develop facial recognition systems — and might easily not have, considering those systems could eventually be used to surveil and recognize them.

While the photographers may have gotten permission to take pictures of these people, some told NBC that the people who were photographed didn’t know their images had been annotated with facial recognition notes and could be used to train algorithms.

“None of the people I photographed had any idea their images were being used in this way,” one photographer told NBC.

The photos weren’t originally compiled by IBM, by the way — they’re part of a larger collection of 99.2 million photos, known as the YFCC100M, which former Flickr owner Yahoo originally put together to conduct research. All the photos were shared under a Creative Commons license, which is typically a signal that they can be freely used, with some limitations.

But the fact they could potentially be used to train facial recognition systems to profile by ethnicity, as one example,

may not be a use that even Creative Commons' most permissive licenses anticipated. It's not entirely a theoretical example: IBM previously made a video analytics product that used body cameras to figure out peoples' races. IBM denied that it would "participate in work involving racial profiling," it tells *The Verge.*

It's also worth noting that IBM's original intentions may have been rooted in preventing AI from being biased against certain groups though — when it announced the collection in January, the company explained that it needed such a large dataset to help train for "fairness" as well as accuracy.

IBM's not alone

Either way, it's hard for the average person to check if their photos were included and request to have them removed, since IBM keeps the dataset private from anyone who's not conducting academic or corporate research. NBC obtained the dataset from a different source and made a tool within its article for photographers to check if their Flickr usernames have been included in IBM's collection. That doesn't necessarily help the people who were photographed, though, if they're not interested in participating.

IBM told *The Verge* in a statement, "We take the privacy of individuals very seriously and have taken great care to comply with privacy principles." It noted that the dataset could only be accessed by verified researchers and only included images that were publicly available. It added, "Individuals can opt-out of this dataset."

IBM is only one of several companies exploring the field of facial recognition and it's not alone in using photos of regular people without expressly asking for their consent. Facebook, for instance, has photos of 800,000 faces open for other researchers to download.

Update March 12th 7:41PM ET: *This article has been updated*

with a statement from IBM.

https://www.theverge.com/2019/3/12/18262646/ibm-didnt-inform-people-when-it-used-their-flickr-photos-for-facial-recognition-training