

The Guardian

Google's solution to accidental algorithmic racism: ban gorillas

Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest



A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

Alex Hern

Fri 12 Jan 2018 11.04 EST

After Google was criticised in 2015 for an image-recognition algorithm that auto-tagged pictures of black people as “gorillas”, the company promised “immediate action” to prevent any repetition of the error.

That action was simply to prevent Google Photos from ever labelling any image as a gorilla, chimpanzee, or monkey - even pictures of the primates themselves.

That's the conclusion drawn by Wired magazine, which tested more than 40,000 images of

animals on the service. Photos accurately tagged images of pandas and poodles, but consistently returned no results for the great apes and monkeys - despite accurately finding baboons, gibbons and orangutans.

Google confirmed that the terms were removed from searches and image tags as a direct result of the 2015 incident, telling the magazine that: "Image labelling technology is still early and unfortunately it's nowhere near perfect".

The gorilla blindness is found in other places across Google's platform: Google Lens, a camera app that identifies objects in images, will also refuse to recognise gorillas. But Google Assistant will correctly identify the primates, as will Google's business-to-business image recognition service Google Cloud Vision.

The failure of the company to develop a more sustainable fix in the following two years highlights the extent to which machine learning technology, which underpins the image recognition feature, is still maturing.

Such technologies are frequently described as a "black box", capable of producing powerful results, but with little ability on the part of their creators to understand exactly how and why they make the decisions they do.

That is particularly true of the first wave of image-recognition systems, of which Google Photos was a part. At the same time that product was launched, Flickr released a similar feature, auto-tagging - which had an almost identical set of problems.

The Yahoo-owned photo sharing platform labelled a picture of a black man as "ape", and a photo of the Dachau concentration camp as "jungle gym". Flickr's response was much the same as Google's: the company apparently removed the word "ape" from its tagging lexicon entirely.

AI programs exhibit racial and gender biases, research reveals **Since you're here ...**

... we have a small favour to ask. More people are reading the Guardian than ever but advertising revenues across the media are falling fast. And unlike many news organisations, we haven't put up a paywall - we want to keep our journalism as open as we can. So you can see why we need to ask for your help. The Guardian's independent, investigative journalism takes a lot of time, money and hard work to produce. But we do it because we believe our perspective matters - because it might well be your perspective, too.

I appreciate there not being a paywall: it is more democratic for the media to be available for all and not a commodity to be purchased by a few. I'm happy to make a contribution so others with less means still have access to information. Thomasine F-R.

If everyone who reads our reporting, who likes it, helps fund it, our future would be much more secure. **For as little as \$1, you can support the Guardian - and it only takes a minute. Thank you.**

Support the Guardian



Topics
Google

/