June 25th, 2018

Mr. Jeffrey P. Bezos
Founder and Chief Executive Officer
Amazon, Inc.
410 Terry Ave. N
Seattle, WA 98109

Re: Audit of Amazon Rekognition Uncovers Gender and Skin-Type Disparities

Dear Mr. Bezos,

As the leader of the Algorithmic Justice League, and a concerned citizen, I join over 150,000 individuals, nearly 70 organizations[1], your 19 concerned shareholders[2], and Amazon employees [3] who have raised their voices in urging you to stop equipping law enforcement with facial analysis technology.

Even if the Amazon Rekognition services and products you are selling to police departments were completely flawless, the potential for abuse on historically marginalized communities would not be reduced.

And as it stands, real-world deployments of facial analysis technology have alarming performance metrics. As I write in the New York Times[4]— " In the case of South Wales, where Big Brother Watch reports[5] that between May 2017 and March 2018 the faces of over 2,400 misidentified innocent people were stored by the police department without their consent, the department reported a false-positive facial identification rate of 91 percent."

After learning of these numbers, I searched for external performance metrics on your technology, and I was unable to find any comprehensive public metrics on the performance of Amazon Rekognition on benchmarks or in the real-world. Thus, on June 20, 2018 I issued an external evaluation on Amazon Rekognition employing the same methodology[6] computer vision expert Dr. Timnit Gebru and I used to test facial analysis technology from IBM, MSFT, and Face++. For binary classification where the chance of obtaining the correct result is 50/50 using a reduced understanding of gender, Amazon Rekognition API performed as follows:

---

[1] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/over-150000-people-tell-amazon-stop-selling-facial

[2] https://www.aclu.org/letter/letter-shareholders-amazon-ceo-jeff-bezos-regarding-rekognition

[3] https://gizmodo.com/amazon-workers-demand-jeff-bezos-cancel-face-recognitio-1827037509

[4] https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html

[5] https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf

[6] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

	All	Females	Males	Darker	Lighter
N	1270	566	704	589	681
Unprocessed Faces	3	0	3	3	0
Accuracy on PF	92.50%	83.39%	99.86%	86.35%	97.80%
Error Rate on PF	7.50%	16.61%	0.14%	13.65%	2.20%

June 20, 2018 Amazon Rekogntion Performance on the Pilot Parlaments Benchmark by Gender and Skin-Type
PF = Processed Faces | Darker = Faces with Fitzpatrick Skin Type IV,V, VI | Lighter = Faces with Fitzpatrick Skin Type I,II, and III

The results are based on performance on the Pilot Parliaments Benchmark, which is a highly constrained dataset of profile photos of African and European legislators who are generally looking directly into a camera in typically well lit conditions. Poor performance on this easy dataset, which does not reflect real-world conditions that present more challenging illumination, pose, and expression variation, suggests it is possible performance in real-world scenarios where individuals may be blurred, not looking directly at the camera, or have their faces partially covered could be even worse. Good performance on this simple benchmark is not a stamp of approval particularly considering the lack of representation of faces from many regions of the world. And critically, improved performance cannot address unfair use of even the most accurate facial analysis technology.

On the easy Pilot Parliaments Benchmark for gender classification, Amazon Rekognition performs better on lighter-skinned faces than darker-skinned faces with an accuracy difference of 11.45 %. It also perform better on male faces than female faces with an accuracy difference of 16.47%. The performance metrics on darker-skinned individuals are especially concerning given the long history of racial bias in policing practices.

To my knowledge Amazon is providing facial identification services for policing which is not the same as gender classification. In the case of gender classification which has been essentially reduced to a binary, the technology has a 1 in 2 chance of getting the answer right simply by guessing. With facial identification, the chance of guessing the correct face by chance is based on the number of potential face matches stored. So for example if there are 50,000 faces to be matched against and a person of interest is identified, the chance of randomly guessing the right individual is 1 in 50,000. And guessing the wrong individual subjects innocents to undue scrutiny as has been reported on by Big Brother Watch.

Given what we know about the biased history and present of policing, the concerning performance metrics of facial analysis technology in real-world pilots, and Rekognition's gender and skin-type accuracy differences on the easy Pilot Parliaments Benchmark, I join the chorus of dissent in calling Amazon to stop equipping law enforcement with facial analysis technology.

Sincerely,

Joy Buolamwini
Algorithmic Justice League