The Washington Post
Democracy Dies in Darkness

TECHNOLOGY

Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use

By Drew Harwell
December 19, 2019 at 6:43 p.m. EST

Facial-recognition systems misidentified people of color more often than white people, a landmark federal study released Thursday shows, casting new doubts on a rapidly expanding investigative technique widely used by law enforcement across the United States.

Asian and African American people were up to 100 times more likely to be misidentified than white men, depending on the particular algorithm and type of search. Native Americans had the highest false-positive rate of all ethnicities, according to the study, which found that systems varied widely in their accuracy.

The faces of African American women were falsely identified more often in the kinds of searches used by police investigators where an image is compared to thousands or millions of others in hopes of identifying a suspect.

Algorithms developed in the United States also showed high error rates for "one-to-one" searches of Asians, African Americans, Native Americans and Pacific Islanders. Such searches are critical to functions including cellphone sign-ons and airport boarding schemes, and errors could make it easier for impostors to gain access to those systems.

Women were more likely to be falsely identified than men, and the elderly and children were more likely to be misidentified than those in other age groups, the study found. Middle-aged white men generally benefited from the highest accuracy rates.

The National Institute of Standards and Technology, the federal laboratory known as NIST that develops standards for new technology, found "empirical evidence" that most of the facial-recognition algorithms exhibit "demographic differentials" that can worsen their accuracy based on a person's age, gender or race.

The study could fundamentally shake one of American law enforcement's fastest-growing tools for identifying criminal suspects and witnesses, which privacy advocates have argued is ushering in a dangerous new wave of government surveillance tools.

The FBI alone has logged more than 390,000 facial-recognition searches of state driver's license records and other federal and local databases since 2011, federal records show. Members of Congress this year have voiced anger over the technology's lack of regulation and its potential for discrimination and abuse.

Lawmakers on Thursday said they were alarmed by the "shocking results" and called on the Trump administration to reassess its plans to expand facial recognition use inside the country and along its borders. Rep. Bennie G. Thompson (D-Miss.), chairman of the Committee on Homeland Security, said the report shows "facial recognition systems are even more unreliable and racially biased than we feared."

Sen. Ron Wyden (D-Ore.) said the findings showed how "algorithms often carry all the biases and failures of human employees, but with even less judgment." In a statement, he added, "Any company or government that deploys new technology has a responsibility to scrutinize their product for bias and discrimination at least as thoroughly as they'd look for bugs in the software."

San Francisco, Oakland and two cities in Massachusetts, Somerville and Brookline, have passed bans this year on facial-recognition use by public officials. The state of California also banned the software's use in police body cameras.

The federal report confirms previous studies from researchers who found similarly staggering error rates. Companies such as Amazon had criticized those studies, saying they reviewed outdated algorithms or used the systems improperly.

One of those researchers, Joy Buolamwini, said the study was a "comprehensive rebuttal" to skeptics of what researchers call "algorithmic bias."

"Differential performance with a factor of up to 100?!?" she wrote The Washington Post in an email Thursday. The study, she added, is "a sobering reminder that facial recognition technology has consequential technical limitations alongside posing threats to civil rights and liberties."

Investigators said they did not know what caused the gap but hoped the findings would, as NIST computer scientist Patrick Grother said in a statement, prove "valuable to policymakers, developers and end users in thinking about the limitations and appropriate use of these algorithms."

Jay Stanley, a senior policy analyst at the American Civil Liberties Union, which sued federal agencies earlier this year for records related to how they use the technology, said the research showed why government leaders should immediately halt its use.

“One false match can lead to missed flights, lengthy interrogations, tense police encounters, false arrests, or worse,” he said. “But the technology’s flaws are only one concern. Face recognition technology — accurate or not — can enable undetectable, persistent, and suspicionless surveillance on an unprecedented scale.”

NIST’s test examined most of the industry’s leading systems, including 189 algorithms voluntarily submitted by 99 companies, academic institutions and other developers. The algorithms form the central building blocks for most of the facial-recognition systems around the world.

The algorithms came from a range of major tech companies and surveillance contractors, including Idemia, Intel, Microsoft, Panasonic, SenseTime and Vigilant Solutions. Notably absent from the list was Amazon, which develops its own software, Rekognition, for sale to local police and federal investigators to help track down suspects.

NIST said Amazon did not submit its algorithm for testing. The company did not immediately offer comment but has said previously that its cloud-based service cannot be easily examined by NIST’s test. Amazon founder and chief executive Jeff Bezos owns The Washington Post.

Grother, the NIST lead researcher, said other companies with cloud-based systems had been able to submit their algorithms, including Microsoft, who he said “sent us very capable and very reliable software.” Of Amazon, he added: “Our test remains open if they elect to participate.”

The NIST team tested the systems with about 18 million photos of more than 8 million people, all of which came from databases run by the State Department, the Department of Homeland Security and the FBI. No photos were taken from social media, video surveillance or the open Internet, they said.

The test studied both how algorithms work on “one-to-one” matching, used for unlocking a phone or verifying a passport, and “one-to-many” matching, used by police to scan for a suspect’s face across a vast set of driver’s license photos. Investigators tested both false negatives, in which the system fails to realize two identical faces are the same, as well as false positives, in which the system identifies two different faces as being the same — a dangerous failure for police, who could end up arresting an innocent person.

Some algorithms produced few errors, but the disparity in accuracy between different systems could be enormous. There is no national regulation or standard for facial-recognition algorithms, and local law-enforcement agencies rely on a wide range of contractors and systems with different capabilities and levels of accuracy. The algorithms themselves — with names such as “anyvision-004” and “didiglobalface-001″ — are almost entirely unknown to anyone outside the industry.

Algorithms developed in Asian countries had smaller differences in error rates between white and Asian faces, suggesting a relationship “between an algorithm’s performance and the data used to train it,” the researchers said.

“You need to know your algorithm, know your data and know your use case,” said Craig Watson, a manager at NIST. “Because that matters.”