Supplementary reading for September 27

Andy Kessler, “Have No Fear of Facial Recognition” Aug. 4, 2019 5:51 pm ET

If it is bound by good legal protections, the technology is a boon, not a tool for tyranny.

Englishman Francis Galton first noted the unique arches, loops and whorls in our fingerprints back in the 1880s. Thirty years later, Clarence Hiller confronted an intruder in his Chicago home and was fatally shot. The culprit fled, but not before leaving a fingerprint in fresh paint on a railing. Thomas Jennings became the first defendant convicted using fingerprints as evidence. This is now routine, but back then there was public hysteria over the fingerprint’s invasion of privacy and then questionable accuracy.

Today, with faces matched almost instantly via machine learning and artificial intelligence, fears of Big Brother have created similar hysteria—especially after Georgetown legal scholars discovered last month that Immigration and Customs Enforcement has access to driver’s license photos from 21 states.

So much so that the crime-ridden cities of San Francisco and Oakland, Calif., along with Somerville, Mass., have banned the use of facial recognition by law enforcement, even though local businesses can use it to track who enters and leaves their buildings. Pretty crazy.

Paranoid? Is someone watching you? Let’s get some constitutional rights out of the way first, especially “unreasonable searches and seizures.” In Katz v. U.S. (1967), the Federal Bureau of Investigation used an electronic eavesdropping device, attached to the outside of a phone booth, to record the defendant’s gambling transactions. Charles Katz won and the Supreme Court ruled that in a phone booth, “like a home, and unlike a field, a person has a constitutionally protected reasonable expectation of privacy.” On the flip side, the justices held that “what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection.”

But we are still protected against dragnets (dum du dum dum)—the use of cameras or other surveillance to track collectively where everyone goes and what they do. Even in public, where you have no reasonable expectation of privacy, law enforcement can’t record everything in hope that someone commits a crime. The USA Patriot Act weakened some of these protections, but the 2015 USA Freedom Act fixed that.

There’s added hysteria around the potential bias in facial recognition’s mistakes. About a year ago, the American Civil Liberties Union did a study, using Amazon’s Rekognition tool, in which it ran photos of members of Congress against a database of 25,000 arrest mug shots. It falsely matched 28 congressmen, 40% of whom were “people of color.” The headlines blared: “Facial recognition’s racial bias problem.”

But if you read the fine print, the ACLU admits that it “used the default match settings that Amazon sets for Rekognition,” which is an 80% confidence level. Amazon reran the study with 30 times as many mug shots and the 99% confidence threshold they recommend for law enforcement use and the “misidentification rate dropped to zero.” You probably missed the media’s retractions.

To see if this technology is any good, I spoke to someone who actually uses it, Capt. Chuck Cohen of the Indiana State Police. He reminisced about the bad old days of passing around grainy videotapes from security cameras asking if “anyone recognized this guy.” He tells me facial recognition is another tool in the forensic shed, along with fingerprints, tire imprints, partial license plates and DNA. He stressed repeatedly that he doesn't consider facial recognition as evidence in court, only as a lead in investigations.

Capt. Cohen says it works. During a physical altercation, someone was shot in the stomach and the victim's friend recorded phone video. Facial recognition identified the shooter, which was the lead police needed to find more evidence to nab the suspect. In another disturbing case, with explicit video, someone sexually harassed and extorted young girls. Police used facial recognition to identify 14 of 22 victims, who were carefully interviewed. They eventually identified the offender.

Hundreds of crimes have been solved. From other sources, I've heard that Alabama security cameras picked up a 90-year-old woman being robbed and beaten by an African-American woman. Police considered lineups, the proverbial usual suspects. Instead they used facial recognition and found the man who did it. You read that right: The culprit was a cross-dressing man, something a lineup would never have found.

The Chinese have powerful facial-recognition tools, SenseTime and Megvii. We know Beijing uses technology for mass surveillance, especially against the Muslim Uighurs. How do we safeguard U.S. citizens against similar abuses? Rather than banning its use, we need strong silos. Such protections exist today. Try getting President Trump's tax returns. Try finding the guy who cut you off with a license-plate number. Cops can't do extended surveillance without a judge's warrant. We can make databases inaccessible except with a judge's consent. Heck, use the judge's face as the ID.

Facial recognition will only get better. But we ought to can the hysteria. So long as the tech is properly limited in use to avoid fishing expeditions, we'll all be safer.

Facial Recognition Tech Is Growing Stronger, Thanks to Your Face

By Cade Metz
July 13, 2019

SAN FRANCISCO — Dozens of databases of people's faces are being compiled without their knowledge by companies and researchers, with many of the images then being shared around the world, in what has become a vast ecosystem fueling the spread of facial recognition technology.

The databases are pulled together with images from social networks, photo websites, dating services like OkCupid and cameras placed in restaurants and on college quads. While there is no precise count of the data sets, privacy activists have pinpointed repositories that were built by Microsoft, Stanford University and others, with one holding over 10 million images while another had more than two million.

The face compilations are being driven by the race to create leading-edge facial recognition systems. This technology learns how to identify people by analyzing as many digital pictures as possible using "neural networks," which are complex mathematical systems that require vast amounts of data to build pattern recognition.

Tech giants like Facebook and Google have most likely amassed the largest face data sets, which they do not distribute, according to research papers. But other companies and universities have widely shared their image troves with researchers, governments and private enterprises in Australia, China, India, Singapore and Switzerland for training artificial intelligence, according to academics, activists and public papers.

Companies and labs have gathered facial images for more than a decade, and the databases are merely one layer to building facial recognition technology. But people often have no idea that their faces are in them. And while names are typically not attached to the photos, individuals can be recognized because each face is unique to a person.

Questions about the data sets are rising because the technologies that they have enabled are now being used in potentially invasive ways. Documents released last Sunday revealed that Immigration and Customs Enforcement officials employed facial recognition technology to scan motorists' photos to identify undocumented immigrants. The F.B.I. also spent more than a decade using such systems to compare driver's license and visa photos against the faces of suspected criminals, according to a Government Accountability Office report last month. On Wednesday, a congressional hearing tackled the government's use of the technology.

There is no oversight of the data sets. Activists and others said they were angered by the possibility that people's likenesses had been used to build ethically questionable technology and that the images could be misused. At least one face database created in the United States was shared with a company in China that has been linked to ethnic profiling of the country's minority Uighur Muslims.

Over the past several weeks, some companies and universities, including Microsoft and Stanford, removed their face data sets from the internet because of privacy concerns. But given that the images were already so well distributed, they are most likely still being used in the United States and elsewhere, researchers and activists said.

"You come to see that these practices are intrusive, and you realize that these companies are not respectful of privacy," said Liz O'Sullivan, who oversaw one of these databases at the artificial intelligence start-up Clarifai. She said she left the New York-based company in January to protest such practices.
"The more ubiquitous facial recognition becomes, the more exposed we all are to being part of the process," said Liz O'Sullivan, a technologist who worked at the artificial intelligence start-up

One database, which dates to 2014, was put together by researchers at Stanford. It was called Brainwash, after a San Francisco cafe of the same name, where the researchers tapped into a camera. Over three days, the camera took more than 10,000 images, which went into the database, the researchers wrote in a 2015 paper. The paper did not address whether cafe patrons knew their images were being taken and used for research. (The cafe has closed.)

The Stanford researchers then shared Brainwash. According to research papers, it was used in China by academics associated with the National University of Defense Technology and Megvii, an artificial intelligence company that The New York Times previously reported has provided surveillance technology for monitoring Uighurs.

The Brainwash data set was removed from its original website last month after Adam Harvey, an activist in Germany who tracks the use of these repositories through a website called MegaPixels, drew attention to it. Links between Brainwash and papers describing work to build A.I. systems at the National University of Defense Technology in China have also been deleted, according to documentation from Mr. Harvey.

Stanford researchers who oversaw Brainwash did not respond to requests for comment. “As part of the research process, Stanford routinely makes research documentation and supporting materials available publicly,” a university official said. “Once research materials are made public, the university does not track their use nor did university officials.”

Duke University researchers also started a database in 2014 using eight cameras on campus to collect images, according to a 2016 paper published as part of the European Conference on Computer Vision. The cameras were denoted with signs, said Carlo Tomasi, the Duke computer science professor who helped create the database. The signs gave a number or email for people to opt out.

The Duke researchers ultimately gathered more than two million video frames with images of over 2,700 people, according to the paper. They also posted the data set, named Duke MTMC, online. It was later cited in myriad documents describing work to train A.I. in the United States, in China, in Japan, in Britain and elsewhere.
Duke University researchers started building a database in 2014 using eight cameras on campus to collect images.

Duke University researchers started building a database in 2014 using eight cameras on campus to collect images. The Duke researchers ultimately gathered more than two million video frames with images of over 2,700 people.
Dr. Tomasi said that his research group did not do face recognition and that the MTMC was unlikely to be useful for such technology because of poor angles and lighting.

“Our data was recorded to develop and test computer algorithms that analyze complex motion in video,” he said. “It happened to be people, but it could have been bicycles, cars, ants, fish, amoebas or elephants.”

At Microsoft, researchers have claimed on the company’s website to have created one of the biggest face data sets. The collection, called MS Celeb, included over 10 million images of more than 100,000 people.

MS Celeb was ostensibly a database of celebrities, whose images are considered fair game because they are public figures. But MS Celeb also brought in photos of privacy and security activists, academics and others, such as Shoshana Zuboff, the author of the book “The Age of Surveillance Capitalism,” according to documentation from Mr. Harvey of the MegaPixels project. MS Celeb was distributed internationally, before being removed this spring after Mr. Harvey and others flagged it.

Kim Zetter, a cybersecurity journalist in San Francisco who has written for Wired and The Intercept, was one of the people who unknowingly became part of the Microsoft data set.

“We’re all just fodder for the development of these surveillance systems,” she said. “The idea that this would be shared with foreign governments and military is just egregious.”

Matt Zeiler, founder and chief executive of Clarifai, the A.I. start-up, said his company had built a face database with images from OkCupid, a dating site. He said Clarifai had access to OkCupid's photos because some of the dating site's founders invested in his company.

He added that he had signed a deal with a large social media company — he declined to disclose which — to use its images in training face recognition models. The social network's terms of service allow for this kind of sharing, he said.

"There has to be some level of trust with tech companies like Clarifai to put powerful technology to good use, and get comfortable with that," he said.

An OkCupid spokeswoman said Clarifai contacted the company in 2014 "about collaborating to determine if they could build unbiased A.I. and facial recognition technology" and that the dating site "did not enter into any commercial agreement then and have no relationship with them now." She did not address whether Clarifai had gained access to OkCupid's photos without its consent.

Clarifai used the images from OkCupid to build a service that could identify the age, sex and race of detected faces, Mr. Zeiler said. The start-up also began working on a tool to collect images from a website called Insecam — short for "insecure camera" — which taps into surveillance cameras in city centers and private spaces without authorization. Clarifai's project was shut down last year after some employees protested and before any images were gathered, he said.

Mr. Zeiler said Clarifai would sell its facial recognition technology to foreign governments, military operations and police departments provided the circumstances were right. It did not make sense to place blanket restrictions on the sale of technology to entire countries, he added.

Ms. O'Sullivan, the former Clarifai technologist, has joined a civil rights and privacy group called the Surveillance Technology Oversight Project. She is now part of a team of researchers building a tool that will let people check whether their image is part of the openly shared face databases.

"You are part of what made the system what it is," she said.