

Massive photo databases secretly gathered in US and Europe to develop facial recognition

Kevin Reed
17 July 2019

A report in the *New York Times* on Sunday revealed that millions of facial photos have been scraped from online sources and taken by hidden surveillance cameras and then shared in databases for artificial intelligence (AI) research and development purposes for more than a decade. Created in secret by universities and tech companies, the photo data sets have been mined for the R&D of facial recognition and biometric technologies that are now used ubiquitously by police and state intelligence agencies around the world.

The large digital face and “selfie” photo databases—copied without authorization from websites, social media, photo sharing and online dating platforms and also taken by digital cameras in public places—have been used by state agencies, software engineers and researchers involved in perfecting AI algorithms and image pattern analyses in the quest for leading-edge facial recognition technology.

According to the *Times* report—based largely on information available on the website MegaPixels.cc published by Adam Harvey and Jules LaPlace—at least 30 facial image datasets were accumulated going back to at least 2007. The *Times* report says that Megapixels “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.”

Summarizing the MegaPixels exposures published online in 2017, the *Times* report went on, “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 ...” Although the *Times* does not mention it, this also includes access to these datasets for testing and development purposes by US government and military agencies through their connections with both the private companies and university research institutions.

For example, a project called Brainwash was launched jointly by Stanford University and the Max Planck Institute for Informatics in Germany in 2014 and deployed a hidden webcam in the Brainwash Café in downtown San Francisco. Stanford University is well known for its connections to US military-intelligence. For example, Google was developed at Stanford with funding from the Defense Advanced Research

Projects Agency (DARPA) and other state intelligence agencies in the early 1990s. Although not mentioned by the *Times*, the Max Planck Institute has long standing and direct ties to German imperialism.

Over a three-day period, 11,917 video streams of 100 seconds each were captured without the consent of those in the Brainwash Café. According to MegaPixels, “No ordinary café customer could ever suspect that their image would end up in dataset used for surveillance research and development, but that is exactly what happened to customers at Brainwash Cafe in San Francisco.”

MegaPixels also said that the videos were published online using AngelCam, a web streaming service that is sold for home security purposes for as little as \$6 per month. The Brainwash database was subsequently used for AI research purposes in China, Switzerland, Netherlands, the US, India and Canada.

In another case, the *Times* reported that Duke University researchers started a facial image database in 2014 called Duke MTMC using eight cameras on campus. The cameras had signs posted below them with a phone number and email address for people who wanted to opt out of the study. Two million synchronized video frames were gathered of approximately 2,700 individuals over 14 hours, most of them students.

However, the *Times* chose to conceal important details regarding US government use of the Duke MTMC dataset. While MegaPixels reports that the Chinese government used the Duke photos—with over 90 research projects in 2018 alone—for surveillance purposes, Harvey and LaPlace also explain that the original creation of the dataset was “supported in part by the United States Army Research Laboratory” and was for “automated analysis of crowds and social gatherings for surveillance and security applications.”

Furthermore, the MegaPixels report says, “Citations from the United States and Europe show a similar trend to that in China, including publicly acknowledged and verified usage of the Duke MTMC dataset supported or carried out by the United States Department of Homeland Security, IARPA, IBM, Microsoft (who has provided surveillance to ICE), and Vision Semantics (who has worked with the UK Ministry of Defence).”

The *Times* also reviewed the Microsoft dataset created in

2016 called MS Celeb that contained 10 million images of 100,000 people gathered from websites that was “ostensibly a database of celebrities.” However, many others had their names and pictures included in the database. Also not mentioned by the *Times* is the fact that MegaPixels published a list of 24 names in the MS Celeb database who are authors, journalists, filmmakers, bloggers and digital rights activists.

Among them is Jeremy Scahill, a journalist and editor with the *Intercept* that has written extensively on US war crimes and defended WikiLeaks editor Julian Assange against imprisonment and rendition to the US. The MS Celeb dataset contains 200 facial photos of Scahill.

The MS Celeb data set had a goal of targeting 1 million people and included an additional 900,000 names that had no images attached. The 100,000-person dataset has been accessed internationally by more than a dozen countries. The MegaPixels web site shows that the MS Celeb data set was cited in 124 research projects that took place around the world in 2018, the majority of which were in China (47) and the US (42).

Two more image databases on the MegaPixels website were not reported by the *Times*, one from Oxford University and the other from University of Colorado. The Oxford Town Centre dataset contains video of 2,200 people captured in 2007 from a surveillance camera mounted at the corner of Cornmarket Street and Market Street in Oxford, England. The surveillance project was commissioned by Oxford University under the auspices of an EU artificial intelligence program called Project HERMES. MegaPixels reports that the image dataset has been shared extensively, with 80 research citations from all over the world.

The final dataset is from the University of Colorado, Colorado Springs campus in which 1,700 students and other pedestrians were “photographed using a long-range high-resolution surveillance camera without their knowledge,” according to MegaPixels. The photos were taken during the spring semester of the 2012-2013 academic year on the West Lawn of the Colorado campus and during the interval that students were walking between classes. MegaPixels reported that the Unconstrained College Student dataset was “providing the researchers with realistic surveillance images to help build face recognition systems for real world applications for defense, intelligence, and commercial partners.”

In total, MegaPixels located 24 million “non-cooperative, non-consensual photos in 30 publicly available face recognition and face analysis datasets” that “were collected without any explicit consent, a type of face image that researchers call ‘in the wild.’ Every image contains at least one face and many photos contain multiple faces. There are approximately 1 million unique identities across all 24 million images.”

Finally, the *Times* reported that a face database was gathered by the software company Clarifai with images from OKCupid, a dating site. Matthew Zeiler, the CEO of Clarifai, told the *Times* that he had access to the OKCupid images because

“some of the dating site’s founders invested in his company.” Zeiler also said that he signed an agreement with a large unnamed social media company “to use its images in training face recognition models.”

Clarifai used the OKCupid photos to develop facial recognition software that can identify the age, sex and race of analyzed faces. When questioned about his intentions by the *Times*, Zeiler said, “Clarifai would sell its facial recognition technology to foreign governments, military operations and police departments provided the circumstances were right.”

The revelation that European- and US-based universities as well as Silicon Valley tech corporations have been involved in gathering “non-cooperative, non-consensual photos” for research purposes for more than ten years shows that the practical implementation of facial recognition and biometrics for state surveillance is well advanced. That these organizations secretly created and shared facial images for AI development also exposes the willingness of significant layers of academia and corporate America to participate overtly in attacking basic democratic rights.

Although the information published in the “independent art and research project” MegaPixels by Adam Harvey and Jules LaPlace—with support from the open source community at Mozilla—has been available since November 2017, the corporate media including the *Times* never saw fit to write about it until now. This is because there is growing public awareness and outrage in the US over facial recognition and biometrics surveillance of the entire population by local, state and federal police agencies.

Additionally, the *Times* story places emphasis on the use of facial image datasets by the Chinese government while deliberately leaving out significant details regarding the role of US, British and German military-intelligence in similar research. This position corresponds to the political and military strategy of ruling factions within these imperialist powers for a more aggressive posture toward China over strategic global interests.

The response of both Democrats and Republicans at every level of government is to push for legislation that will establish a legal framework for using facial recognition and AI tools to spy on the people. It is to this objective that the latest reports from the *Times* are directed and this is why certain key facts—especially those regarding the role of US military-intelligence—have been excluded from their coverage.



To contact the WSWS and the Socialist Equality Party visit:

wsws.org/contact