ELECTRONIC
FRONTIER
FOUNDATION

January 27, 2020

VIA MAIL

Purdue University
School of Electrical and Computer Engineering
465 Northwestern Avenue
West Lafayette, IN 47907-2035

RE: NIST TATT-C DATASET

To Whom It May Concern:

I am writing today on behalf of the Electronic Frontier Foundation (EFF), a non-profit public interest organization that uses advocacy, litigation, and computer science to protect civil liberties as technology advances. Since 2015, EFF has been investigating a series of tattoo recognition projects managed by the National Institute of Standards and Technology (NIST) and the Federal Bureau of Investigation (FBI). Our research—and subsequent Freedom of Information Act (FOIA) litigation—revealed multiple problems in the ethical process and failures to protect the privacy of people whose images were distributed as part of the projects.

You are receiving this letter because your organization requested and is believed to have received access to a dataset of tattoo images as part of the Tatt-C challenge, according to documents obtained through our FOIA litigation. This letter details the many ethical flaws of the Tatt-C program and the privacy and free expression problems that arise in attempts to create automated tattoo recognition tools.

In light of these problems, EFF requests that your organization take immediate action to address the series of privacy abuses and ethical lapses associated with this dataset.

These actions must include:

1) The destruction of the Tatt-C dataset;
2) The initiation of an internal review of all research generated using the Tatt-C dataset;
3) The initiation of a review of the institution's policies for training biometric recognition algorithms using images or other biometric data collected from individuals who did not consent to be photographed and did not consent to the images being used to train algorithms.

Please confirm the receipt of this letter as soon as possible.

We request that your institution respond by February 21, 2020 with information regarding the action it will take regarding items 1-3.

We intend to publish a list of entities that have taken appropriate action and those that have yet to address the ethical challenges associated with the Tatt-C dataset. We will publicize the status of each entity on a regular basis until such time as each entity that received the data has confirmed deletion and initiation of the reviews.

Should you have further questions, please email Dave Maass at dm@eff.org. He may also be reached by phone at +1 415 436-9333, extension 151.

Tatt-C Dataset Overview

NIST's Image Group launched its Tattoo Recognition Technology program in 2014, with the assistance of the FBI's Biometric Center for Excellence.[1] The program's stated goal is to "advance research and development into automated image-based tattoo recognition technology."

The program framework laid out a series of projects, the first of which was the Tattoo Recognition Technology Challenge (Tatt-C).[2] For Tatt-C, NIST and FBI compiled an "open tattoo database" of approximately 15,000 images collected from prisoners.[3] NIST offered the data to universities, research institutions, and companies to conduct tattoo recognition research through the following process: parties would file a request with NIST. Then the parties would sign a second agreement with the FBI, which would mail the dataset on a CD-ROM to the requester.

NIST intended for the dataset to be the first standardized metric for testing tattoo recognition algorithms. Entities that wanted to participate in the challenge were instructed to self-report the accuracy of their tattoo recognition algorithms. However, participation in the challenge was not required in order to access the data; researchers were encouraged to use the dataset for their own experiments and several requested the data after the Tatt-C testing had ended.[4]

[1] NIST Tattoo Recognition Program page: https://www.nist.gov/programs-projects/tattoo-recognition-technology

[2] NIST Tattoo Recognition Challenge page: https://www.nist.gov/programs-projects/tattoo-recognition-technology-challenge-tatt-c

[3] Mei Ngan and Patrick Grother, *Tattoo Recognition Technology - Challenge (Tatt-C): An Open Tattoo Database for Developing Tattoo Recognition Research*, International Conference on Identity, Security and Behavior Analysis (ISBA), pp.1-6 (2015), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=917896

[4] Mei Ngan, George W. Quinn and Patrick Grother, *Tattoo Recognition Technology – Challenge (Tatt-C) Outcomes and Recommendations*, NIST Interagency Report, 26 (Sept. 2016) https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8078.pdf

Records released to EFF through Freedom of Information Act litigation indicate that nearly 30 institutions asked for and received the Tatt-C dataset by filing a formal request with NIST and the FBI.[5]

Tatt-C concluded in September 2015 and was followed by the Tattoo Recognition Technology Evaluation project (Tatt-E).[6] This research used a larger tattoo dataset--also largely collected from prisoners—but the experiments were conducted by NIST internally and did not involve the release of further data.

Dataset Did Not Receive Human Subjects Evaluation

When compiling the Tatt-C dataset and conducting the Tatt-C challenge, NIST researchers failed to follow the Common Rule's requirements for ethical review of the project. Specifically, the researchers did not seek a Human Subjects determination, which is required before assessing whether research must be reviewed by an independent review board (IRB). According to official reports, "the human subjects determination was not known to be necessary for the proposed use of data."[7]

NIST discovered the oversight only after the Tatt-C research was conducted. At that stage, Information Technology Lab Director Charles H. Romine took the unusual step of retroactively determining that the research did not involve human subjects.[8] EFF's evaluation of the application and determination found multiple red flags, including erroneous claims that the data was sufficiently de-identified from the subjects and downplaying NIST's role in creating the data set. Many of the images reviewed by EFF contained personally identifying information, including people's names, faces, and birth dates.[9] Of great concern was the researchers' omission that the data was collected from prisoners, which generally triggers greater scrutiny.

It is likely your organization acted on the good-faith belief that the Tatt-C dataset had been properly vetted by the U.S. government and moved through the appropriate ethical review process. But the documents we have obtained show that this was not the case, and NIST's failure to adequately evaluate ethical concerns before the creation of this database of prisoners' photos cast shadows on the entire research process.

[5] *EFF v. Dep't of Commerce*, No. 17-cv-02567 (D.D.C. 2017), documents produced: https://assets.documentcloud.org/documents/6148072/17-Cv-02567-4-2019-NIST-Referral.pdf

[6] Tattoo Recognition Evaluation page: https://www.nist.gov/programs-projects/tattoo-recognition-technology-evaluation-tatt-e

[7] Charles H. Romine, Determination for ITL Project #ITL-0002, entitled, "Tattoo Recognition Technology Research and Evaluations with Federal Bureau of Investigation (FBI) Operationally Collected Tattoo Data," (April 10, 2015), https://www.eff.org/document/nist-tattoo-recognition-foia

[8] Ibid.

[9] Aaron Mackey and Dave Maass, *Tattoo Recognition Research Threatens Free Speech and Privacy*, EFF Deeplinks (June 2, 2016), https://www.eff.org/deeplinks/2016/06/tattoo-recognition-research-threatens-free-speech-and-privacy.

Dataset Contained PII

When the Tatt-C test corpus of approximately 15,000 tattoo images was distributed to entities, the FBI represented that "This data has been cleansed of all Personally Identifiable Information (PII)."[10] This representation has proven to be erroneous.

While NIST claimed that the images in the Tatt-C were coded in such a way "insure that the identities of the contributing subjects cannot be readily ascertained," it failed to take into account the uniqueness of the tattoos and that the very purpose of biometric recognition is to identify an individual.

While the FBI and NIST may have deleted PII from the metadata and indices for each image, the agencies failed to assess the images themselves for PII. As a result, many of the images included PII such as photo-realistic images of relatives, the names of relatives, and dates of birth and deaths of relatives. In one example, EFF took information contained in the tattoo and was able to identify the individual within minutes with a Google search.

After EFF raised concerns about PII included in the images themselves, NIST retroactively removed images containing PII and other potential identifiers from the dataset in its possession and began omitting most images from the related materials.[11] However, it did not have the ability to delete the PII from the datasets already in possession of third parties.

NIST and FBI's evaluation of the dataset also failed to consider the potential for reidentification of images when combined with other datasets.

While a person in a photograph may not have their name attached to a particular image in the data set, those images are still unique to the person and can be matched to other datasets that do reveal an individual's name, such as datasets compiled from Flickr or other social media sites. Documents produced in response to our FOIA suit include a presentation showing that researchers at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation had the ability to match tattoos from websites to a national criminal database.[12] Researchers at Nanyang Technological University used the Flickr API to download thousands of images, which it then used in research that also involved the NIST dataset.[13]

[10] Documents released through the Freedom of Information Act: https://assets.documentcloud.org/documents/6148072/17-Cv-02567-4-2019-NIST-Referral.pdf (Page 9 of PDF)

[11] Dave Maass, *Documents Bare How Federal Researchers Went to Absurd Lengths to Undo Problematic Tattoo Recognition Research*, EFF Deeplinks (August 21, 2018), https://www.eff.org/deeplinks/2018/08/eff-bares-how-federal-researchers-went-absurd-lengths-undo-problematic-tattoo

[12] Dave Maass, *Researchers Matched Images on Tattoo Websites to a German Police Database*, EFF Deeplinks (Nov. 3, 2016), https://www.eff.org/deeplinks/2016/11/researchers-matched-images-tattoo-websites-german-police-database

[13] Soham Ghosh, *Tattoo Detection Based on CNN and Remarks on the NIST Database*, Nanyang Technological University (June 15, 2016), https://web.archive.org/web/20190912205137/https://pdfs.semanticscholar.org/82ad/be4e36027f90994a7fc7249c67b54839d36b.pdf

In EFF's view, had the existence of PII been surfaced in application materials, as well as the involvement of prisoner specimens, NIST should have reached another conclusion in its human subjects determination. It should have classified the program as human subjects research, triggering an IRB review. Had this process been followed, EFF believes it is likely that an IRB may have rejected the overall endeavor due to the enhanced standards required for research involving prisoner subjects.

Tattoo Research Implicates Human Rights and Civil Liberties

A. Expression issues

Unlike faces or fingerprints, tattoos are not simply a physical characteristic: they are an expression of identity. When a person makes a choice to get a tattoo, they are engaging in speech. Whether that's a tattoo promoting their favorite sports team, celebrating the birth of a child, or a traditional tattoo tied to one's heritage, it's rare for a tattoo not to be an expression of the wearer's culture and beliefs. In recognizing the First Amendment right to get a tattoo, and limitations on the government from preventing citizens from expressing this right, the Ninth Circuit Court of Appeals has said, "We have little difficulty recognizing that a tattoo is a form of pure expression entitled to full constitutional protection."[14]

Tattoos may further implicate a number of related rights enshrined in the First Amendment of the Constitution. Tattoos of crosses have been an expression of Coptic Christianity for more than 700 years.[15] A member of a labor union may get a tattoo to express their freedom to associate and organize.[16] A number of journalists are tattooed with images embracing their pride in the trade.[17]

Even NIST has recognized the speech implications of tattoos, and has acknowledged that the ability to identify, track, and group speech is a benefit to law enforcement of this research. In a Tatt-C presentation before participants in the research, NIST justified the usefulness of the tattoo recognition challenge saying tattoos "suggest affiliation to gangs, subcultures, religious or ritualistic beliefs, or political ideology."[18] When NIST originally published its research paper on the Tatt-C in September 2015, the second sentence of the paper's Introduction explicitly stated: "Tattoos provide

[14] *Anderson v. City of Hermosa Beach*, 621 F.3d 1051 (9th Cir. 2010).

[15] Adelaide Mena, *Holy tattoo! A 700-year old Christian tradition thrives in Jerusalem*, Catholic News Agency (July 9, 2017), https://www.catholicnewsagency.com/news/holy-tattoo-a-700-year-old-christian-tradition-thrives-in-jerusalem-68723

[16] Diane S. Williams, *Organizer wears labor pride on his sleeve*, Public Employee Press (June 2018), https://www.dc37.net/news/PEP/6_2018/labor_pride

[17] Examples:
Former New York Times Editor Jill Abramson's tattoo: https://www.huffpost.com/entry/jill-abramson-tattoo-new-york-times_n_5155322
Washington Post Data Editor Stephen Rich's tattoo: https://twitter.com/dataeditor/status/307674739994918913?lang=en
Wonkette Founder Ana Marie Cox's tattoo: https://twitter.com/anamariecox/status/436172526637813760

[18] Aaron Mackey and Dave Maass, *Tattoo Recognition Research Threatens Free Speech and Privacy*, EFF Deeplinks (June 2, 2016), https://www.eff.org/deeplinks/2016/06/tattoo-recognition-research-threatens-free-speech-and-privacy.

valuable information on an individual's affiliations or beliefs and can support identity verification of an individual."[19] In a section titled "Application scenarios," NIST described "Tattoos have utility in supporting group affiliation such as membership to street/prison gangs and terrorist/hate groups as well as threat assessment on suspects."[20] These claims underlined the portion of the NIST project described as "Tattoo Similarity," in which algorithms were tasked with identifying tattoos on different people, but with similar imagery. Two of the nine images displayed in the research paper involved iconography associated with Catholicism: praying hands with a rosary and Jesus during the crucifixion.[21]

A year later, following criticism in the press, NIST altered its materials to retroactively delete these references from its published materials. The new version erases the troubling nature of the original research, without making any disclosure in the release notes about what was removed and why.

At best, NIST's original framing of the Tatt-C challenge was tone deaf. At worst, it was a gross violation of the subjects' rights to freedom of expression, religion, and association that may lead to the development of government tools that could be used to persecute or discriminate against individuals based on their religion, beliefs, affiliations, and ethnicity.

B. Prisoners as a Commodity

The NIST tattoo projects also disregard prisoners' dignity and basic human rights, treating their tattoos--often captured from areas of the body normally covered by clothing--as a commodity. Prisoners did not have an opportunity to consent or refuse participation or provide comment on the program. And yet, NIST provided their sensitive images to third parties--such as your institution--to hone tattoo-recognition algorithms. This software in turn may be licensed or sold through lucrative deals with government agencies. Indeed, MorphoTrak issued a press release promoting its success in the Tatt-C trials as a reason public safety agencies should use its products.[22]

Research like Tatt-C exploits prisoners at little to no cost and without the lengthy consent process. This is deeply troubling in light of the historic origins of ethical principles such as the Common Rule, which were created in part to prevent unethical experimentation involving prisoners. As noted above, NIST did not follow the process for IRB review for prisoner-involved research, and instead retroactively declared that the prisoner images were not human subject specimens.

[19] Mei Ngan, George W. Quinn and Patrick Grother, *Tattoo Recognition Technology – Challenge (Tatt-C) Outcomes and Recommendations*, NIST Interagency Report 8078 (Sept. 2015), https://www.eff.org/files/2019/12/05/tatt-c_outcomes_and_recommendations.pdf

[20] Ibid.

[21] Ibid.

[22] Sharon Rollins, *MorphoTrak Tattoo ID First in NIST Evaluation* (August 4, 2015), https://web.archive.org/web/20170630191108/https://www.morpho.com/en/media/morphotrak-tattoo-id-first-nist-evaluation-20150804

Current Ethical Process Must Meet Contemporary Practices

Over the last five years, a number of scandals involving biometric datasets have revealed that existing ethical frameworks for research are not suitable for research involving biometric recognition. In 2019, Microsoft deleted the MSCeleb database of 10-million images used for facial recognition testing after it was revealed that the image set included many individuals engaged in sensitive activities, such as journalists and human rights activists.[23] Ethical questions were also raised about a University of Colorado, Colorado Springs facial recognition dataset which was compiled using a camera taking surreptitious portraits of students on campus.[24] Although the professor behind this "Unconstrained Students" dataset defended the project, he acknowledged that the set improperly captured the date and time the images were taken, which "thwarted the intended purpose of trying to randomize the photos in the dataset." Duke University conducted similar data collection of students using surreptitious photograph. In this case, Duke removed the dataset after an internal investigation found the images were "neither collected nor made available to the public consistent with the terms of the study that had been approved by the Institutional Review Board."[25] Critics of these programs have noted that these datasets, such as Microsoft and Duke, may have been used by face recognition researchers in authoritarian nations such as China.

When NIST's tattoo recognition experiments were re-evaluated by the agency's human subjects reviewer, she made an unusual determination that the prisoner images were not "subjects" of the experiment. The rationale was that because the project was measuring the efficacy of algorithms that the algorithms were the "subjects" not the human specimens. As a result, NIST absolved itself of the responsibility of evaluating whether the program could harm inmates.

NIST's conclusion is troubling. It allows researchers to sidestep the ethical issues in underlying data that identifies people or contains their images so long as the end result of the research is an algorithm derived from the data. Taken to its logical conclusion, NIST's rationale would allow any research aimed at designing or improving algorithms to avoid ethical scrutiny even when the underlying data is taken from humans, including vulnerable groups like prisoners.

[23] *Microsoft deletes massive face recognition database*, BBC News (June 7, 2019), https://www.bbc.com/news/technology-48555149

[24] Elizabeth Hernandez*, CU Colorado Springs students secretly photographed for government-backed facial-recognition research*, Denver Post (May 27, 2019), https://www.denverpost.com/2019/05/27/cu-colorado-springs-facial-recognition-research/

[25] Jake Satisky, *A Duke study recorded thousands of students' faces. Now they're being used all over the world*, The Duke Chronicle (June 11, 2019), "https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur

Remediation

In light of the many problems with NIST's dataset and research described above, we believe it is imperative that your institution take immediate steps to rectify the harms caused by the Tattoo Recognition experiments.

To address this, first and foremost, your institution must delete all copies of the Tatt-C dataset immediately. Second, your institution should initiate an investigation into all tattoo recognition research to ensure that such research is privacy protective, ethical, and does not infringe on human rights and free speech. Finally, we call on your institution to review its ethical framework to ensure that the protections in place for subjects is adequate to protect the rights of those whose biometrics are captured, stored, shared, and analyzed. Any policy should address issues of consent and evaluate the potential harms that the research itself might cause as well as the distribution of datasets compiled as part of the experiments. It should also include strong protections to ensure that researchers cannot use images from prisoners without following the guidelines long-established to ethically and sensitively engage prisoners in research.

Please provide a written response by February 21, 2020. This response should include information regarding the actions you will take to remedy the issue and confirm that the dataset has been deleted. EFF intends to publish your involvement in this program, including your response, as a part of a public campaign intended to establish ethical norms for biometric research.

Should you have any questions, do not hesitate to contact Dave Maass at dm@eff.org or by phone at 415-436-9333 x151.

Sincerely,

Dave Maass
Senior Investigative Researcher
Electronic Frontier Foundation