Maintaining American Values with AI Ethics Standards for Data Collection and Algorithmic Processing

Introduction

This comment is written in response to NIST's request for comments on the creation of

> "*AI technical standards and tools in support of reliable, robust and trustworthy systems that use AI technologies*".[1]

Specifically, this comment is written in response to the original Executive Order, which states that

> *"the US must foster public trust and confidence in AI technologies and protect civil liberties, privacy and American values in their application in order to fully realize the potential of AI technologies for the American people*."[2]

In both statements of the request, an emphasis is placed on making standards that enable the American people to trust any AI technology that is created. The current collection of public comments addresses this idea of trust by focusing on technical issues. Whether it's trying to avoid tampering from foreign countries, ensuring high quality training data, or trying to democratize access to data systems, these approaches all operate on the fundamental belief that "a trustworthy system is a technically correct system".

Although these are admirable concerns, there is a critical component missing from these responses – developing trust by respecting the consent and privacy of the American people. Trust is not simply a matter of technical prowess; it comes from respecting the civil liberties of the people who will be affected by AI systems. Thus, although NIST has not traditionally considered more social elements into their standards, the sheer impact that these AI systems have on society means that any standards NIST issues must consider how to ensure proper consent and usage of data and algorithmic processing.

Concern over Existing NIST Standards

This focus on the social impact of trust is not without precedent. Chinese oppression of the Uighur minority population in Xinjiang has been built largely through the rise of AI-enabled technologies[3] . Facial recognition technologies allow institutionalized racial profiling, while app-enabled monitoring limit movement and freedom of assembly [4]. If the purely technical lens to trust is applied to these technologies, nothing would seem to be wrong. These systems work fine at their task, and we can trust that they adequately classify people or track people across the country. However, the social lens to trust is broken as these technologies should not have been created or deployed in the first place.

---

[1] https://www.federalregister.gov/d/2019-08818/p-16
[2] https://www.federalregister.gov/d/2019-08818/p-9
[3] https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html
[4] https://www.nytimes.com/2019/02/21/business/china-xinjiang-uighur-dna-thermo-fisher.html

This problem is not just limited to Chinese government, as American companies and scientific research have also occasionally been (unintentionally) complicit in this surveillance state. American-created DNA databases and commercial genetic sequencing tools have been used without informed consent to create massive genetic surveillance [5]. If NIST wishes to follow the mandate from the Executive Order and truly create standards that enable trustworthy technologies, standards must be written to ensure that similar technologies are not created here in the US.

However, NIST's existing facial recognition datasets cause great concern for the potential of future AI technologies due to the dubious nature of their subjects' consent [6]. NIST's Face Recognition Vendor Test (FRVT) sets a standard for government agencies, researchers and industry on facial recognition technologies [7]. However, by compiling images from other government agencies, such as the Department of Homeland Security, FRVT contains images of people who did not consent to being used as test subjects for training AI systems. In particular, these images range from children exploited for pornography[8], US visa applicants[9], and dead people who have committed multiple crimes[10]. None of the people imaged are probably aware of their inclusion in datasets, let alone given explicit consent.

In particular, the second evaluation report of the FRVT[11] occasionally highlights that permission was granted for some images (ex. Fig. 6 and 7). This further underscores how permission was not granted or solicited in cases of mugshots and driver licenses (Fig. 2, 3, 9). The report also uses the non-consensual acquisition of images as a positive attribute for the diversity of the dataset, highlighting "Accuracy with non-cooperating subjects" in Appendix E.

Although these datasets may have been compiled with appropriate approval for their single use case, there has been no follow-up to determine whether all other future uses of these data do as well, especially since consent continues to be unsolicited. Even if these databases improve technical accuracy of AI systems, FRVT's current focus on marginalized populations as test cases is uncomfortably close to China's focus on the Uighur population for racial profiling.

Recommendations for Future Standards

In order to fulfill the mandate of "fostering public trust and confidence in AI technologies", NIST must ensure that its standards and its existing datasets provide respect for the autonomy of the people which the data is derived from. To achieve this goal, the author provides the following recommendations:

[5] https://www.nytimes.com/2019/02/21/business/china-xinjiang-uighur-dna-thermo-fisher.html
[6] https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html
[7] https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt
[8] https://www.nist.gov/programs-projects/chexia-face-recognition
[9] https://www.hsdl.org/?view&did=438078
[10] https://www.nist.gov/itl/iad/image-group/special-database-32-multiple-encounter-dataset-meds
[11] https://doi.org/10.6028/NIST.IR.8238

(1) AI systems should actively acknowledge and solicit consent from not only the people that explicitly build and design the system, but also the "invisible labor" of the people used to train the system – whether it's the marginalized groups who form the basis of these datasets or the people who processed all of the child pornography images to highlight facial features.

(2) The standards of data justice / design justice should be incorporated into any standards that NIST creates, in order to ensure appropriate buy-in.[12] Part of these principles involve designing technical systems in conversation with the end user as well as ensuring greater transparency about the data processing techniques. This will result in standards that not only say how datasets should be used, but also how they *shouldn't* be used.

(3) NIST should recommend against a "release-and-forget" model of dataset creation. Datasets cannot simply be moved from one venue to another just because consent was given in one domain. Legal scholars have already noted that traditional frameworks of "deidentification" and "anonymization" are fundamentally broken, as all data inherently becomes personally identifiable, especially in the face of improved AI technologies. [13]

The author applauds NIST's solicitation for feedback on future AI standards. She hopes that ethics will make it into future AI standards and that these ethical standards can also be applied to NIST's previous datasets as well.

Sincerely,

Lillian Chin

PhD Student, Electrical Engineering and Computer Science
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

May 31, 2019

This response is the opinion of the author only and does not necessarily represent the opinion of MIT as an institution.

[12] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3189696
[13] https://www.uclalawreview.org/pdf/57-6-3.pdf