CENTER FOR
DEMOCRACY
& TECHNOLOGY

Comments of the Center for Democracy & Technology

in response to

The National Institute of Standards and Technology

Request for Information:

Developing a Federal AI Standards Engagement Plan

June 10, 2019

The Center for Democracy & Technology (CDT) thanks the National Institute of Standards and Technology for the opportunity to comment on its request for information regarding artificial intelligence, standards, and NIST's role in developing a national strategy for AI.[1] CDT is a non-partisan, non-profit organization working to preserve civil rights and democratic ideals online and in existing and new applications of technology. As such, CDT is interested in the federal government's approach to emerging technologies and supports NIST's efforts to address standardization with respect to artificial intelligence.[2]

NIST should play a central role in the government's comprehensive efforts to address AI. Standardization of some facets of AI could benefit the development of the technology itself, while standardization of the ways agencies approach new technologies stands to benefit both the government and the public. Even beyond the scope of the Executive Order, CDT encourages NIST to pursue its efforts in this space.[3] CDT looks forward to future engagements with NIST and offers the expertise of its staff and that of the GRAIL Network wherever it can be helpful.[4] Note that throughout this comment we use the terms *artificial intelligence* (AI) and *machine learning* (ML) somewhat interchangeably, reflecting the fact that more general forms of artificial intelligence are not particularly ripe for standardization.

More generally, CDT notes that some commenters suggest proprietary standards to which access is restricted.[5] CDT suggests that unrestricted access to government-supported standards is preferable because it promotes greater transparency and accountability for those implementing the standards and

[1] U.S. Dep't. Of Commerce, National Institute for Standards and Technology, *Request for Information: Artificial Intelligence Standards*, 84 Fed. Reg. 18490, (May 1, 2019), at 18490-92.

[2] CDT thanks Erik Stallman and Rob Walker of the Samuelson Law, Technology, and Public Policy Clinic at the University of California, Berkeley Law School, for their guidance and expertise.

[3] *Executive Order on Maintaining American Leadership on Artificial Intelligence*, (Feb. 11, 2019), https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

[4] The Governance and Research in Artificial Intelligence Leadership Network is a project of CDT and the R Street Institute. GRAIL consists of leading experts from the academic and research communities studying AI, from the computer and data science foundations to the public policy implications, and is intended to facilitate more informed policy discussions by helping researchers engage with policy makers.

[5] *See, e.g.*, Comments of IEEE at 3, proposing P7001 as a standard that "describes measurable, testable levels of transparency, so that autonomous systems can be objectively assessed."

for those agencies tasked with oversight and assessment of standard compliance. Furthermore, CDT suggests that NIST is ideally situated to conduct open standard setting processes, in which a broader diversity of interests may be addressed more so than what may result from industry-led efforts at standardization.

In response to the request for information, CDT respectfully submits the following comments:

Privacy, Security, and Transparency Considerations for Data Sets

Data is the raw material input into ML systems. Although neither data sets nor the models derived from them are uniform in structure, content, format, purpose, or use, there are some aspects of data for which standardization would be beneficial. CDT suggests that greater uniformity in the methods for handling and securing data and in the ways we talk about data would be worthy of NIST's further consideration.[6]

With the Big Data Public Working Group project, NIST has already begun to address some opportunities for uniformity with respect to the storing, securing, and processing of large data sets.[7] For example, NIST's "Unlinkable Data Challenge" offers an incentive to create more robust methods for removing connections between data sets and individual people.[8] Those methods may then form the basis for information processing standards and help to protect billions of people from unwanted identification or association. Once NIST has selected the most effective methods from the Challenge, it should consider how to encourage broader deployment of the best techniques, including by people and companies using data sets to develop ML models and other AI-related products. In short, NIST has already taken some important steps toward standardization for data practices; it should now apply that work to this new context.

In addition to privacy and security practices for data sets, information about data and data sets, or metadata, (including traditional metadata like file size or time of creation as well as more *meta*-metadata like data provenance, set size, etc), is perhaps as critical to automated systems as the underlying information. Metadata helps everyone from data scientists to policy makers understand context like where the data came from, the degree to which it fairly represents the subject, and the purposes for which the data would be appropriate. NIST has already proposed a schema for evaluating attribute metadata, but there are other features of data and data sets for which standardized metadata

[6] CDT generally agrees with commenters noting that federally-created data sets may be useful for both development and testing of systems. See, e.g. comments of IEEE at 10.

[7] Wo L. Chang, Arnab Roy, Mark Underwood, NBD-PWG NIST Big Data Public Working Group, *NIST Big Data Interoperability Framework: Vol. 4, Big Data Security and Privacy (Version 2)*, (June 26, 2018), https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-4-big-data-security-and-privacy-version

[8] NIST, *Help Keep Big Data Safe by Entering NIST's 'Unlinkable Data Challenge'*, (May 1, 2018), https://www.nist.gov/news-events/news/2018/05/help-keep-big-data-safe-entering-nists-unlinkable-data-challenge

conventions could be useful.[9] For instance, information about data sets, rather than attribute values, would help developers and policy makers understand and compare the quality, representativeness, uses, and limitations of the data sets as well as the ML models derived from them.

The concept of “datasheets for data sets” provides a framework for data set creators to provide documentation about much of this metadata, such as the intended use cases of the dataset and necessary maintenance, as well as known biases in the data set. This would facilitate better evaluation of ML systems and their training data, as well as promote transparency and stronger accountability for developers.[10] Datasheets would help developers understand the strengths and limitations of datasets, while giving auditors, purchasers, and users of ML systems a tool for assessing vendors’ claims about their products and a consistent rubric for comparing the foundational elements of ML systems. NIST is well situated to lead or coordinate the federal government’s efforts with respect to standardizing information formats for data sets, whether by proposing a standard format for dataset metadata reporting, improving on existing formats, or coordinating with other agencies to develop a datasheet format for federal datasets.

NIST seems equally well situated to identify any differences among federal agencies’ treatment, handling, presentation of, and reporting on their data sets. Identifying these inconsistencies or gaps in standardization for government data practices would allow better prioritization and resource allocation toward consistent government practices. Additionally, to the extent that NIST and other agencies can align any inconsistent practices, doing so could improve the government’s ability to leverage its data sets across multiple agencies and data applications.

Recommendations: Pursue greater uniformity in the methods for handling and securing data, standardize methods for conveying information about data sets, and identify inconsistent approaches to handling and describing data at the federal level.

ML Systems: Structures, Elements, and Vocabulary

It is useful to think of a ML technique as encompassing three distinct technical aspects: the type of underlying model, the evaluation method between model outputs, and optimization performed across potential outputs.[11] The type of model—or, equivalently the representation of the classifier—corresponds to the basic approach the model uses to “learn,” for example, techniques such as K-nearest neighbor, decision trees, neural networks, and genetic algorithms. Model outputs must be

[9] Paul A. Grassi Naomi B. Lefkovitz Ellen M. Nadeau Ryan J. Galluzzo Abhiraj T. Dinh, *Attribute Metadata A Proposed Schema for Evaluating Federated Attributes*, (January 2018), NIST https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8112.pdf

[10] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford, *Datasheets for Datasets*, (2018), available at: https://www.fatml.org/media/documents/datasheets_for_datasets.pdf

[11] Pedro M. Domingos, *A few useful things to know about machine learning*, *Comm. ACM* 55, no. 10 (2012): 78-87, available at: https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf.

evaluated to understand how well the model is working. A variety of evaluation statistics can be used (and in many cases, more than one in concert), including accuracy, precision and recall, squared error, and likelihood estimation. Finally, to search through a potential forest of evaluated model outputs, it is necessary to use an optimization function—often a form of grid or random search, but increasingly much more efficient and complex methods such as gradient descent—which aim to find the best candidate models for the problem at hand.

Different types of ML techniques have different kinds of configuration parameters and hyperparameters—parameters that are set before any training or learning commences, such as the number of layers in a neural network or the learning rate in a gradient descent optimization. Standardizing aspects of ML "flavors" are important to create a standard vocabulary with which different ML researchers can discuss how they tune their algorithms. In addition, different configurations of ML systems require different kinds of controls and parameters that are not well-standardized now, such as ways of creating testing or training data sets for supervised learning systems.[12]

Recommendations: establish a uniform vocabulary for describing structures, elements, parameters, hyperparameters, and techniques for developing ML systems.

Evaluating and Comparing System Performance

Whether for the purposes of procurement or regulation, the federal government would benefit from using a standardized process and set of criteria by which to compare and assess various AI systems. However, the diversity among potential applications of AI is so broad that not all systems will be responsive to the same technical criteria.

Evaluation in ML itself can be exceedingly complicated and has mostly focused on task-based and peer-vs-peer evaluation to-date, with some indications that less anthropocentric and more general, ability-focused types of evaluation methods are on the horizon.[13] The older work of Cohen and Howe lays out a number of useful considerations that can be generally applied to AI problems to promote evaluation and comparison supplementing NIST's rich experience with hosting benchmark data sets and competitions in this space.[14] [15]

[12] For example, validation of an ML model can be performed via "holdout" methods where a subset of data is retained for testing (which can result in high variance) versus forms cross-validation where models are trained on many "withheld" subsets that are each used to test the model (but can result in model over-fitting). *See* Prashant Gupta, *Cross-Validation in Machine Learning*, Towards Data Science (June 5. 2017), available at: https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f.

[13] José Hernández-Orallo, *Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement*, Artificial Intelligence Review 48, no. 3 (2017): 397-447.

[14] Paul R. Cohen and Adele E. Howe, *How evaluation guides AI research: The message still counts more than the medium*, AI magazine 9, no. 4 (1988): 35-35. Note: at the time of writing of this work in 1988, academic computer science was not nearly as focused on evaluation as it is now, so some of their recommendations must be evaluated themselves in light of modern practices and norms.

CDT suggests that NIST could work toward developing a broadly applicable framework that agencies could use to help understand systems from a policy perspective, such as how systems are alike, how they differ from one another, whether and how they are subject to existing regulations, and how well they might work to solve a particular problem. For example, NIST could further develop the concept of "model cards," which is similar to the "datasheets for datasets" idea in that they would help to clarify the capabilities and limitations of any statistical models used in ML applications.[16]

In the context of procurement, agencies will need a consistent process by which to compare systems to other similar systems and to evaluate how well those systems match the needs of the agency. Such comparisons may be difficult if vendors do not use the same artifacts, benchmarks, or metrics to describe the performance of their products. NIST could help agencies as they begin to add more AI systems to their portfolios by creating standards for evaluating systems' capabilities and qualities.

Similarly, as agencies look ahead to the regulation of AI systems, it will be necessary to identify which systems are subject to regulatory oversight and to systematically assess system functions with respect to regulatory requirements, such as transparency and access mandates. A standard framework or process for evaluating systems would improve consistency for agencies and provide greater certainty for regulated entities.[17]

A consistent approach to evaluating and comparing AI systems would help purchasers, regulators, and policy makers by helping to ensure that they compare "apples to apples" and to assess whether a system provides the best solution for a given problem. CDT encourages NIST to develop standards for performing these assessments.

Recommendations: develop a broadly applicable framework to enable policy makers to understand, evaluate, and compare AI systems.

Reproducibility and Replicability

Being able to reproduce results from AI and ML systems and replicate findings based on new data is a critically important aspect of promoting research, development, and innovation in this area. There is an active effort in the scientific community to build reproducibility and replicability into the practice of

[15] Including Open Machine Translation (OpenMT), Text Analysis Conference (TAC), Performance Metrics for Intelligent Systems Workshops (PerMIS), and special reference data such as the EMNIST handwritten character database.

[16] While model cards could help to understand some ML systems, they might not be applicable to other types of systems, such as neural nets. *See* Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T., *Model cards for model reporting*, (January 2019), Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229). ACM, available at: https://arxiv.org/abs/1810.03993

[17] Although not addressed here, the speed of ML systems may be relevant in some scenarios. One example of a speed benchmarking project is MLPerf, which demonstrates the kind of structured, standardized comparisons that might be useful in other contexts. See https://mlperf.org.

scientific inquiry. The National Academies of Science, Engineering, and Medicine has recently published, "Reproducibility and Replicability in Science,"[18] which defines concepts such as computational reproducibility and replicability as:

> ***Reproducibility** is obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis. This definition is synonymous with "computational reproducibility," [...] **Replicability** is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.*

Stodden and Miguez effectively break down reproducibility as having three essential components: "1) what do we need to capture from the original computational environment? 2) how can [that] information be communicated to [other researchers] and be both functional and persistent over time? [and] 3) what are the sources of errors or uncertainty, even given the same data, code, and environment?"[19] While there are good practices, frameworks, and tools for designing reproducibility into research and development efforts, they could use more generic standardization. NIST appears well qualified to do so.

Unfortunately, replicability is a much more complicated issue, often requiring deep, time-consuming and expensive investigation into potential sources of disagreement when the same method and computational tools are applied to different collections of data.[20] Here, it might be more important for NIST to convene researchers working in AI and ML to understand how standards might help improve replicability or facilitate identifying sources of non-replicability.

Recommendations: convene stakeholders to work toward standardization of good practices, frameworks, and tools for designing reproducibility and support for replicability into research and development efforts.

Auditing: Evaluating Systems Against Legal and Social Norms

Most AI systems, especially those designed to address complex problems, should be seen as experimental designs that must be closely monitored. While much can be done during the development process to address potential sources of bias, inaccuracy, or ineffectiveness in automated systems, no AI can be assumed to remain free of these issues once put into use. These systems require regular auditing to ensure that they are accurate, fair, predictable, and in compliance with legal and ethical

[18] National Academies of Sciences, Engineering, and Medicine, *Reproducibility and Replicability in Science* (2019) Washington, DC: The National Academies Press. https://doi.org/10.17226/25303. (NASEM report)
[19] Victoria Stodden and Sheila Miguez, *Provisioning Reproducible Computational Science*, (2014), available at: https://www.xsede.org/documents/659353/703287/xsede14_stodden.pdf.
[20] *See* NASEM report, at 71-86.

standards.[21] Internal or independent evaluators must be able to examine components of an automated system, such as its training data, source code, and inputs and outputs. Robust auditing is necessary to "foster trust and confidence in AI technologies and protect civil liberties, privacy and American values[,]" a chief objective of the Executive Order on AI.[22]

Just as a range of field experiments is used to detect bias and discrimination in social settings, there is a range of software auditing techniques that thoughtful and consistent standards can support.[23] The level of visibility into the automated system will often dictate the autiting method. Some methods are designed to function in the face of extremely limited visibility, and "treat [the system's] decision process as a black box whose inputs and outputs are visible but the inner workings are unseen."[24] The American Civil Liberties Union's (ACLU) recent test of Amazon's facial recognition tool, Rekognition, is one example of this "black box" auditing. There, the ACLU used Rekognition to search a database of 25,000 publicly available arrest photos against photos of every current member of the U.S. House and Senate, resulting in 28 false matches.[25]

In addition to scrutinizing inputs and outputs from an automated system, auditing techniques may also examine the datasets used to train an AI system, its source code, and the weights given to distinct variables in a system's statistical models. While none of these give perfect information about how an AI functions, all of them yield valuable insights into potential sources of bias or other flaws.

For example, researchers have analyzed the data used to train predictive policing algorithms and concluded that automated systems trained on racially biased arrest records will replicate this bias in their operations.[26] The conclusion is straightforward: biased training data leads to biased results. To reach such a conclusion, however, evaluators need some insight into the composition and operation of automated systems beyond what black-box testing may provide—in this case, information about the training data used. More visibility into AI systems enables more forms of independent evaluation that in turn yield more insights into how systems make decisions and the potential biases or inaccuracies that may creep into the decision making process.

[21] CDT agrees with other commenters noting that standardizing verification and validation process could lead to better auditability for systems designed and built with those standards in mind. See, e.g. Comment of IEEE, at 10. CDT also agrees with commenters observing that auditing should be an ongoing process. *See Id*.

[22] Executive Order on Maintaining American Leadership on Artificial Intelligence § 1(d), Feb. 11, 2019, https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

[23] *See* Devah Prager, *The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future*, 609 Annal of the Am. Acad. of Political and Social Science, 104, 108–14 (discussing variations on correspondence tests and in-person audits to investigate employment discrimination).

[24] Joshua Kroll et. al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633, 660–61 (2017).

[25] Jacob Snow, *Amazon's Face Recognition Falsely Matched 20 Members of Congress With Mugshots*, ACLU, July 26, 2018, https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28[https://perma.cc/3Z38-4L98].

[26] Kristian Lum and William Isaac, *To Predict and Serve?*, 13 Significance 14, 16 (2016) (discussing a predictive policing algorithm used by the Oakland Police Department).

To this end, standards for transparency in the development and use of AI systems can help provide those insights, particularly if they keep independent evaluation in mind. The growing discussion and interest in enhanced and uniform disclosures—whether "datasheets for datasets," a "data statement schema," "dataset nutrition labels," "model cards," or "a supplier's declaration of conformity"—are very encouraging developments.[27] However, any standard will more effectively promote trust and confidence in AI if, ex ante, the development stage of AI systems (which in many cases is never complete) contemplates and provides for the ex post and independent auditing of those systems. As discussed above, standardized presentation of data, metadata, models, developer goals, and the impact of automated systems can improve development, explainability, comparability, and auditability of those systems. For example, describing clearly both "the data bias policies that were checked" against the development of a model or system as well as ongoing "bias checking methods[] and results" may assist independent researchers in performing their own verifications of those policies.[28]

Consistent standards need not be inflexible ones. Indeed, standards should keep pace with new developments in methods for independent investigation of automated systems. Alongside black-box testing, "sock puppet audits," and other tried-and-true auditing methods, researchers are investigating how to apply advanced computer science techniques that support accountability of an AI system "even when the software or the data input to it is secret[.]"[29] [30] These emerging methods are better viewed as supplements to transparency than a replacement for it, but standards should encourage both.

In particular, NIST could collaborate with agencies such as the Department of Housing and Urban Development (HUD) and the Equal Employment Opportunity Commission (EEOC) to advance auditing for disparate impacts across race, gender, and other protected characteristics. Academic and journalistic investigations have revealed evidence of AI systems' capacity to facilitate discrimination in housing, financial, and employment opportunities.[31] While some companies may be internally

[27] M. Arnold et. al., *FactSheets: Increasing Trust in AI Services Through Supplier's Declarations of Conformity*, Feb. 7, 2019, https://arxiv.org/pdf/1808.07261.pdf.
[28] *Id*. at 28.
[29] Christian Sandvig et. al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* 13, May 22, 2014, https://perma.cc/V8ED-R83M (presented at 64th Annual Meeting of the International Communications Association). ("A sock puppet audit is essentially a classic audit study but instead of hiring actors representing different positions on a randomized manipulation as 'testers,' the researchers would use computer programs to impersonate users, likely by creating false user accounts or programmatically-constructed traffic.")
[30] Joshua Kroll, *supra* note 23 at 662–71 (discussing software verification, cryptographic commitments, zero-knowledge proofs, and fair random choices).
[31] *See, e.g.*, Till Speicher et al., *Potential for Discrimination in Online Targeted Advertising*, Proceedings of Machine Learning Research 81:1–15, 8, T. 2 (2018), http://proceedings.mlr.press/v81/speicher18a/speicher18a.pdf; Julia Angwin, Ariana Tobin & Madeleine Varner, *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, ProPublica (Nov. 21, 2017), https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin; Julia Angwin, Noam Scheiber & Ariana Tobin, *Dozens of Companies are Using Facebook to Exclude Older Workers from Job Ads*, ProPublica (Dec. 20, 2017), https://www.propublica.org/article/facebook-ads-age-discrimination-targeting; Amit Datta, Michael Carl Tschantz &

developing methods for disparate impact testing, few share substantive information (either publicly or privately among companies) about effective methods or promising experiments.

One roadblock is the challenge of testing for disparate impact without explicitly collecting sensitive characteristics such as race. In some regulated fields where disparate impact testing and reporting are required, such as in credit and lending, proxy models have been used to infer race.[32] However, these inference methods can face accuracy problems.[33] Uninterpretable ML models or limited feedback data can also inhibit model owners' views into how their systems are impacting different groups. Despite these challenges, disparate impact testing is critical for detecting and mitigating racial and other disparities in AI, ensuring fair and trustworthy systems, and complying with civil rights law.[34]

Recommendations: build standards for transparency in the development and use of AI systems; provide research, guidance, and standardization for disparate impact testing methods, including methodologies for inferring sensitive characteristics; develop data security and privacy considerations for collecting and retaining sensitive testing data; and standardization of disparate impact testing datasets that can be shared among different entities.

Learning from Existing Regulatory Activities

AI and ML are being deployed by a wide variety of industries, many of which are already governed by regulatory frameworks aimed at preventing at least some of the harms which could come from misuses of these technologies. In response, regulators are already beginning to recognize and develop responses to common problems with AI. As NIST works to help other federal agencies address the new

Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, In Proceedings on Privacy Enhancing Technologies (2015), https://arxiv.org/abs/1408.6491; Amit Datta et al., *Discrimination in Online Advertising: A Multidisciplinary Inquiry*, in Proceedings of Machine Learning Research 81:1–15, 3–7 (2018), http://proceedings.mlr.press/v81/datta18a/datta18a.pdf; Muhammad Ali et al., *Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes* (Apr. 19, 2019), https://arxiv.org/pdf/1904.02095.pdf. CDT agrees with commenters suggesting that AI systems should be examined for harms to underrepresented, vulnerable, and marginalized populations" before deployment. Comments of Partnership on AI. CDT is a member of the Partnership on AI.

[32] *See* Consumer Financial Protection Bureau, *Using publicly available information to proxy for unidentified race and ethnicity: a methodology and assessment*, (2014) https://www.consumerfinance.gov/data-research/research-reports/usingpublicly-available-information-to-proxy-for-unidentified-race-and-ethnicity/.

[33] Jiahao Chen et al., *Fairness Under Unawareness: Assessing Disparity When Protected Class is Unobserved*, In FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19) at 29–31, 2019, https://dl.acm.org/authorize.cfm?key=N675485.

[34] CDT notes that commenters proposed specific standards in response to this need. Comments of IEEE, at 4. However, CDT neither supports nor objects to this standard other than to reiterate that publicly accessible standards improve transparency and accountability for companies and regulatory agencies.

demands of AI, we encourage the agency to consider, learn from, and where appropriate, replicate the policies, analytical frameworks, and tools developed by other agencies.[35]

CDT and many others have identified a series of concerns regarding AI and ML. Three of the most noteworthy for this context are:

- Disparate impact - models that unfairly disadvantage individuals based on protected characteristics;
- Accountability - models which make decisions that cannot be easily explained or understood. This is critical for redress (a person can only dispute the accuracy of a decision if she knows what information and reasoning went into making that decision) and fairness (a person cannot seek to achieve a better outcome in a future decision if she does not understand the factors which contributed to the current decision); and
- Reliability - models that fail to meet their stated goals. This is particularly significant for high stakes decisions such as the provision of employment, credit or housing.[36]

While ML and AI may exacerbate these problems or present them in new ways, the problems themselves are not new. Rather, it is well understood in a variety of areas that we do not want decisions to be biased, unaccountable, or wrong. More importantly, existing law already bars it.

For example, the Equal Credit Opportunity Act (ECOA) prohibits discrimination both in the treatment of an applicant and through the use of seemingly neutral policies or practices that have an adverse impact on a prohibited class unless the policy or practice serves a legitimate business need which cannot be achieved in a less discriminatory way.[37] Similarly, when credit is denied, ECOA requires that the applicant receive an adverse action notice which provides “the applicant with the specific principal reason for the action taken.”[38] The Consumer Financial Protection Bureau makes clear that these cannot be generic notes: “[s]tatements that the adverse action was based on the creditor’s internal standards or policies or that the applicant, joint applicant, or similar party failed to achieve a qualifying score on the creditor’s credit scoring system are insufficient.”[39]

Other statutes have similar prohibitions. The Fair Credit Reporting Act requires an adverse action notice for any use of a report to “deny your application for credit, insurance, or employment” and grants

[35] One potential model of a consistent approach to agencies’ self-evaluation of AI systems would be that of impact assessments. See Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, (April 2018), https://ainowinstitute.org/aiareport2018.pdf.

[36] See CDT, *Digital Decisions*, https://cdt.org/issue/privacy-data/digital-decisions/

[37] Consumer Financial Protection Bureau, *Consumer Laws and Regulations: Equal Credit Opportunity Act*, (June 2013), https://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf

[38] 12 C.F.R. § 1002.9(a)(2).

[39] *Id.*

consumers the right to dispute information that is “inaccurate, incomplete, or unverifiable.”[40] The Fair Housing Act prohibits advertisements “which deny a particular segment of the housing market information about housing opportunities because of race, color, religion, sex, handicap, familial status, or national origin.”[41]

In each of these areas, AI technologies have either already been deployed or there are immediate, near-term plans to use the technology.[42] For example, according to a recent media report, beginning in the second half of this year ZestFinance will partner with Discover to use ML and AI to help make underwriting decisions for personal loans:

> *Hundreds of data points gathered from loan applications and elsewhere will be considered in the underwriting process, according to people familiar with the matter. Among those who will be viewed with more suspicion: applicants who claim income in the low six figures or higher—well beyond the median for U.S. households—and those who list an employer’s full legal name, a possible sign a swindler is copying and pasting information, the people said.*
>
> *On the other hand, applicants who call Discover from a landline or cellphone, rather than Skype or other internet-phone services, will be considered safer bets because they’re easier to trace back to an individual, the people added.*[43]

ZestFinance has said these new types of credit decisions, which rely on alternative data, will be driven by ML “by taking advantage of interpretable machine learning approaches to make more accurate lending decisions.”[44] Credit reporting agencies have also expressed interest in making more active use of AI.[45]

The use of these types of alternative credit are becoming more commonplace. As the Treasury Department notes in a recent report, “the use of data for credit underwriting is a core element of online marketplace lending, and one of the sources of innovation that holds the most promise and risk. While data-driven algorithms may expedite credit assessments and reduce costs, they also carry the risk of

[40] Consumer Financial Protection Bureau, *A Summary of Your Rights Under the Fair Credit Reporting Act*, https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf.
[41] 24 C.F.R. § 100.75(c)(3)
[42] *See, e.g.*, Miranda Bogen & Aaron Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*, UpTurn, (December 2018), https://www.upturn.org/reports/2018/hiring-algorithms/. *See also*, Bonnie G. Buchanan, *Artificial Intelligence in Finance*, Alan Turing Institute, (March 27, 2019), available at https://zenodo.org/record/2612537.
[43] AnnaMaria Andriotis, *Shopping at Discount Stores Could Help Get You a Loan*, Wall Street Journal, (March 4, 2019), https://www.wsj.com/articles/use-a-landline-that-could-help-you-get-a-loan-from-discover-11551695400.
[44] Zest Finance Team, *Discover Partners With Zest To Improve Its Credit Underwriting With AI*, (March 4, 2019), https://www.zestfinance.com/blog/discover-partners-with-zest-to-improve-its-credit-underwriting-with-ai.
[45] Alan Ikemura, *Machine learning for real-world credit risk*, Experian, (Sept. 12, 2018) http://www.experian.com/blogs/insights/2018/09/machine-learning-real-world-credit-risk/

disparate impact in credit outcomes and the potential for fair lending violations."[46] In another example, a major advertiser, Facebook, has been accused of relying on algorithms which use prohibited characteristics to shape what housing advertisements individual users were shown.[47]

As ML and AI play a key role in a wide variety of legally cognizable decisions, agencies will have to grapple with the application of existing regulatory requirements to these new decisions. CDT believes that NIST can play a vital role in this process by identifying how the agencies are overseeing entities subject to these requirements and highlighting those lessons for broader regulation and engagement on AI. As standards are developed for bias assessment, explainability, and reliability, NIST can integrate them into its existing efforts to promote conformity in standards.

Recommendations: consider, learn from, or replicate the policies, analytical frameworks and tools developed by other agencies for applying existing regulatory requirements to AI and ML systems.

Respectfully submitted,

Stan Adams

Chris Calabrese

Natasha Duarte

Joseph Lorenzo Hall

Hannah Quay-de la Vallee

Center for Democracy & Technology

1401 K Street, NW Suite 200

Washington, DC 20005

(202) 637-9800

June 10, 2019

[46] U.S. Dep't of the Treasury, *Opportunities and Challenges in Online Marketplace Lending*, (May 10, 2016), https://www.treasury.gov/connect/blog/Documents/Opportunities_and_Challenges_in_Online_Marketplace_Lending_white_paper.pdf

[47] *U.S. Dep't of Housing & Urban Development v. Facebook, Inc.*, FHEO No. 01-18-0323-8, (March 28, 2019), https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf