Federal Register Notice 86 FR 46278, https://www.federalregister.gov/documents/2021/08/18/2021-17737/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence, October 1, 2021.

Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views and/or opinions of the U.S. Government nor those of the National AI Research Resource Task Force., and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality, or content of all external links included in this document.

ACLU
AMERICAN CIVIL LIBERTIES UNION

National Office
125 Broad Street
18th Floor
New York, NY 10004
aclu.org

Deborah N. Archer
President

Anthony D. Romero
Executive Director

October 1, 2021

VIA EMAIL

Office of Science & Technology Policy /
National Science Foundation

RE: Request for Information on an Implementation Plan for a National Artificial Intelligence Research Resource (86 Fed. Reg. 39081)

I. Introduction

While there may be value in creating a shared resource where American researchers can build and develop artificial intelligence ("AI") tools for many different applications, it would be a mistake to assume that investment in AI is inherently beneficial. The current framing of this resource takes the technochauvinist[1] view that the most pressing problem with AI is its inaccessibility and that the democratization of AI will inevitably lead to positive returns for the United States. But expansions of AI that lack an express, ongoing focus on how AI development will affect civil rights and civil liberties will invariably lead to technologies that threaten these important protections. Applications of AI are already used in decisions that fundamentally impact people's lives, ranging from who gets held in jail to who gets a job, loan, or insurance. The discrimination in these uses of AI continues unabated. Regulation and legislation have not kept pace, failing to identify and root out AI bias or halt applications of AI where the risk of discrimination is too great. People are increasingly becoming aware of the ways AI and algorithmic decision-making systems may already be negatively affecting them, causing a degree of "techlash."[2] Current questions around AI abound: when does prediction lead to discrimination; when do the goals of increased safety and efficiency creep into unwelcome surveillance; and in what ways does current AI innovation trap us in past assumptions, dooming the future to repeat the past? These questions must be grappled with and addressed from the outset as part of any initiative to encourage AI development.

II. Lack of Civil Society Representation on the Task Force and Ways the Task Force Can Constructively Engage Going Forward

Harms to the individuals and communities upon whom AI is ultimately deployed can originate at every stage of the AI lifecycle—in problem and domain selection,

[1] Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* 7–8 (2018) ("Technochauvinism is the belief that tech is always the solution.").
[2] *See The Techlash Against Amazon, Facebook, and Google—and What They Can Do*, The Economist (Jan. 20, 2018), https://www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do [https://perma.cc/Z8YK-D5EL].

in data collection and selection, and in ideation, development, and deployment of AI. And these harms occur in critical areas impacting individual freedom as well as social and economic opportunity. Myriad forms of bias and harm have already been identified in the collection and use of data and deployment of AI in the criminal legal system,[3] housing,[4] employment,[5] credit,[6] and education.[7]

The NAIRR Task Force has a critical role to play in determining the path that AI development will take, and robust consideration of these potential harms must be front and center in the Task Force's work. As discussed in more detail below, this includes decisions about ethical standards for whether or not to make certain data sets available in the first place, whether or not to support the development of particular AI tools, and how to identify and prevent uses of the NAIRR that would create unacceptable harms to individuals and communities. Further, there must be transparency and accountability in decision-making around what datasets are made available and the technologies that are ultimately deployed through use of those datasets.[8] Additionally, in setting "parameters for the establishment and sustainment of the National Artificial Intelligence Research Resource, including agency roles and responsibilities,"[9] the NAIRR Task Force should (1) create specific and concrete agency goals, responsibilities, and milestones regarding substantive discussions to address a variety of bias, equity, transparency, accountability, and civil rights/liberties issues ("harms") that may arise in the development, execution and maintenance of the NAIRR and (2) establish concrete ways to meaningfully identify and address these harms in the development, execution, and maintenance of the NAIRR.

The primary impediment we see to creation of a NAIRR that consistently and appropriately reinforces principles of ethical and responsible research, development, and deployment of AI is

---

[3] Michelle Bao et al., *It's COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks* (June 10, 2021), https://arxiv.org/abs/2106.05498 [https://perma.cc/H44H-QBAG].

[4] Letter from ACLU et al. on Addressing Technology's Role in Housing Discrimination to U.S. Dep't of Hous. & Urb. Dev. et al. (July 13, 2021), https://www.aclu.org/letter/coalition-memo-re-addressing-technologys-role-housing-discrimination [https://perma.cc/52RV-PNBF]; Patrick Sisson, *Housing Discrimination Goes High Tech*, Curbed (Dec. 17, 2019), https://archive.curbed.com/2019/12/17/21026311/mortgage-apartment-housing-algorithm-discrimination [https://perma.cc/M3XH-B3PD].

[5] Letter from ACLU et al. on Addressing Technology's Role in Hiring Discrimination to U.S. Equal Emp. Opportunity Comm'n et al. (July 13, 2021), https://www.aclu.org/letter/coalition-memo-addressing-technologys-role-hiring-discrimination [https://perma.cc/W39Q-SQKT]; Ctr. for Democracy & Tech., Algorithm-driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination? (2020), https://cdt.org/wp-content/uploads/2020/12/Full-Text-Algorithm-driven-Hiring-Tools-Innovative-Recruitment-or-Expedited-Disability-Discrimination.pdf [https://perma.cc/9A74-WCCF].

[6] Letter from ACLU et al. on Addressing Technology's Role in Financial Services Discrimination to Consumer Fin. Prot. Bureau et al. (July 13, 2021), https://www.aclu.org/letter/2020-07-13-coalition-memo-technology-and-financial-services-discrimination [https://perma.cc/58YA-4YSQ]; Letter from Nat'l Fair Hous. All. on Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning to Fed. Rsrv. Sys. et al. (July 1, 2021), https://nationalfairhousing.org/wp-content/uploads/2021/07/Federal-Banking-Regulator-RFI-re-AI_Advocate-Letter_FINAL_2021-07-01.pdf [https://perma.cc/H529-V84L].

[7] Ctr. for Democracy & Tech., *Algorithmic Systems in Education: Incorporating Equity and Fairness When Using Student Data* (Aug. 12, 2019), https://cdt.org/insights/algorithmic-systems-in-education-incorporating-equity-and-fairness-when-using-student-data/ [https://perma.cc/PKY6-K6DR].

[8] *See* Crystal Grant & Kath Xu, *Public Trust in Artificial Intelligence Starts with Institutional Reform*, ACLU (Sept. 17, 2021), https://www.aclu.org/news/national-security/public-trust-in-artificial-intelligence-starts-with-institutional-reform [https://perma.cc/MJP4-6VKC].

[9] Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource, 86 Fed. Reg. 39081 (July 23, 2021).

that the NAIRR Task Force itself lacks sufficient expertise in the array of bias, equity, transparency, accountability, and civil rights/liberties issues that can flow from AI.[10] Indeed, given the current structure of the Task Force and plan for its work, we have grave concerns that these harms are unlikely to be sufficiently addressed in the research, development, execution, and administration of the NAIRR.

First, while the members of the NAIRR Task Force and the panelists invited to appear before it to date are distinguished and qualified in many respects, they lack depth and range of expertise with civil rights/ liberties issues. This is not for a lack of such expertise in the field of AI. Indeed, there are many technical experts who have deeply engaged with how AI interplays with systemic discrimination on the basis of gender, race, disability or other protected characteristics, and yet such expertise is not sufficiently represented among the members of the Task Force.[11] Without including experts in bias, equity, transparency, accountability, and civil rights/liberties on the Task Force, these considerations are likely to be sidelined and ultimately viewed as an impediment to the development of AI. Furthermore, absent specific expertise on potential harms, an accurate and holistic "assessment of, and recommend[ed] solutions to, barriers to the dissemination and use of high-quality government data sets" is not possible.[12]

Second, the structure of and plan for the working groups is insufficient to adequately identify potential civil rights/liberties harms and create a plan to concretely address them in the development, execution, governance, and administration of the NAIRR. The Task Force's work plan for considering these harms appears limited to the work of a separate "privacy, civil rights, and civil liberties" working group that only convenes in the final stage of the NAIRR Task Force's assessment phase.[13] Structural bias in big data and the harms it can cause to humans, particularly vulnerable or marginalized communities, cannot be an afterthought. Instead, the Task Force should embed consideration of bias, equity, transparency, accountability, and civil rights/liberties in its "goals and evaluations metrics" as well as in each of the working groups. In particular, the working groups on "governance models," "data resources," "user interfaces," "educational tools," and

[10] *See* Letter from ACLU et al. to White House Off. of Sci. & Tech. Pol'y (July 13, 2021), https://www.aclu.org/sites/default/files/field_document/2021-07-13_letter_to_white_house_ostp_on_centering_civil_rights_in_ai_policy_1.pdf [https://perma.cc/B3F8-Q29Q].

[11] Margaret Mitchell et al., *Diversity and Inclusion Metrics in Subset Selection*, 2020 Proc. AAAI/ACM Conf. on AI, Ethics, and Soc'y (Feb. 7–8, 2020), https://dl.acm.org/doi/pdf/10.1145/3375627.3375832 [https://perma.cc/3J5W-TJDT]; Margaret Mitchell et al., *Model Cards for Model Reporting*, 2019 Conf. on Fairness, Accountability & Transparency (Jan. 29-31, 2019), https://arxiv.org/pdf/1810.03993.pdf?source=post_page [https://perma.cc/539H-S4QW]; Inioluwa Deborah Raji et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing*, 2020 Conf. on Fairness, Accountability & Transparency (Jan. 27–30, 2020), https://dl.acm.org/doi/pdf/10.1145/3351095.3372873 [https://perma.cc/W43W-BHNU]; Joy Buolamwini & Timnit Gebru, Gender Shades: *Intersectional Accuracy Disparities in Commercial Gender Classification*, 2018 Conf. on Fairness, Accountability & Transparency (Feb. 23–24, 2018), https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification/ [https://perma.cc/343P-NRV9]; Deborah I. Raji & Jingying Yang, *ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles*, 33rd Conf. on Neural Info. Processing Sys. (2019), https://arxiv.org/pdf/1912.06166.pdf [https://perma.cc/K5EQ-DLKM]; Timnit Gebru et al., *Datasheets for Datasets* (Mar. 19, 2020) (preprint), https://arxiv.org/abs/1803.09010v7 [https://perma.cc/NL72-R75K].

[12] S. 3890, 116th Cong., 2d Sess. (2020), https://www.congress.gov/116/bills/s3890/BILLS-116s3890is.xml [https://perma.cc/3ZQJ-48X5].

[13] Presentation, Nat. AI Rsch. Res. Task Force (Aug. 30, 2021), https://www.ai.gov/wp-content/uploads/2021/09/NAIRR-TF-Presentations-08302021.pdf [https://perma.cc/3WN3-JJBD].

"testing resources" must include robust consideration of these harms and ways to identify and address them. Preferably, each working group would include at least one member with substantive expertise in these harms so that their consideration can be incorporated in all the working groups' recommendations.

Third, nothing in the Task Force's publicly available agendas, minutes, and presentations suggests that there is a concrete plan for meaningful engagement with individuals and communities who have either experienced harm in the past from collection of data and deployment of AI, or who might suffer AI-related harms in the future. Although the Task Force appears to have heard from some experts on the topic of equity in access to AI education and funding, the Task Force also needs to include robust discussions about the potential harms that the NAIRR may pose to individuals and communities from whom data is collected and upon whom AI is deployed. End-users and impacted communities are often distinct and, unfortunately, the latter group is frequently left out of discussions about the use of data and AI. Each working group should create plans for how they will meaningfully solicit input on these topics from a variety of impacted communities and other experts on an ongoing basis. Presentation at a single meeting or brief consultations with experts will not provide for the deep considerations of bias, equity, transparency, accountability, and civil rights/liberties issues that is required to understand potential negative impacts on people and communities. To foster "public trust" and "prioritize areas of AI that solve problems for the public good"—issues that the panel on "Defining the Value Proposition and Intended Outcomes of a NAIRR" flagged as important[14]—it is critical that representatives of communities that may be impacted by the NAIRR, as well as other experts in the aforementioned harms, take part in the governance and decision-making about the NAIRR.

Several people and institutions possess both the type of technical expertise contemplated by the National AI Initiative Act, as well as expertise in the racial, gender, disability, and other biases that can arise in the collection and use of data and in development and deployment of AI. The Task Force could consider including civil rights/liberties organizations with particular expertise in data and AI, such as those listed below, in its working groups, list of consultants, and as candidates for participation in the governance structure of the NAIRR. Suggested organizations include but are not limited to: AI Now (https://ainowinstitute.org/), Algorithmic Justice League (https://www.ajl.org/), American Civil Liberties Union (https://aclu.org/), Artificial Intelligence, Policy, and Practice Initiative at Cornell University (https://aipp.cis.cornell.edu/), Center for Democracy and Technology (https://cdt.org/), Data 4 Black Lives (https://d4bl.org/), Georgetown Law School's Center for Privacy and Technology (https://www.law.georgetown.edu/privacy-technology-center/), the Lawyers' Committee for Civil Rights Under Law (https://www.lawyerscommittee.org/), the Leadership Conference (https://civilrights.org/), and Upturn (https://upturn.org/).

The Task Force should also solicit input from federal agencies that are tasked with working on bias, equity, and civil rights/liberties, including: the Civil Rights Division of the Department of Justice, Consumer Federal Protection Bureau, Department of Housing and Urban Development, Department of Labor, Equal Employment Opportunity Commission, Federal Trade Commission, and Office of Federal Contract Compliance Programs.

[14] Meeting Summary, Nat'l AI Rsch. Res. Task Force (Aug. 30, 2021), https://www.ai.gov/wp-content/uploads/2021/09/NAIRR-TF-Meeting-Minutes-08302021.pdf [https://perma.cc/AE2W-9CUQ].

III. Task Force Emphasis on Large-Scale Computing Resources Reifies a Type of AI That Is Especially Costly for Society

We believe that the Task Force has, to date, over-emphasized the role of large-scale compute in its considerations. This leads to the myopic view that building bigger and more expansive compute is necessary for positive advances in the AI field. The Task Force needs to consider recent scholarship questioning the tradeoffs associated with building larger models as it decides where to invest limited government resources.[15] The financial[16] and environmental[17] costs for small gains in accuracy are substantial in the places where increasing computational power works—and in many domains (e.g., tabular data), adding compute or ever more complex models may not do much to improve performance.[18] The hoovering up of ever-larger datasets turns them into "stochastic parrots" of some of the most toxic information within that data[19] and can lead to increasingly exploitative data collection practices.[20] The private sector[21] and the government itself[22] have now invested many billions in capital in ever more expansive AI, yet many of the most vexing problems remain, and little of this investment addresses the harms these models may inflict on society. The Task Force should consider whether investing in the compute-race is the best use of its limited resources. Some of the biggest proponents of the compute-heavy AI that the Task Force envisions are now questioning the emphasis on these models,[23] and the Task Force should as well.

---

[15] Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, Proc. of the 2021 ACM Conf. on Fairness, Accountability, and Transparency 610 (2021), https://dl.acm.org/doi/pdf/10.1145/3442188.3445922 [https://perma.cc/MD4V-EYAN] (critiquing large language models); Song Han, Huizi Mao, & William J. Dally, *Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding*, Int'l Conf. on Learning Representations (2016), https://arxiv.org/abs/1510.00149 [https://perma.cc/8YL9-JAHN] (focusing on high efficiency as opposed to ever-larger models).

[16] *See The Staggering Cost of Training SOTA AI Models*, Synced (June 27, 2019), https://medium.com/syncedreview/the-staggering-cost-of-training-sota-ai-models-e329e80fa82 [https://perma.cc/3D9Q-G77F].

[17] *See* Mark Labbe & Ronald Schmelzer, *AI and Climate Change: The Mixed Impact of Machine Learning*, EnterpriseAI (Aug. 31, 2021), https://searchenterpriseai.techtarget.com/feature/AI-and-climate-change-The-mixed-impact-of-machine-learning [https://perma.cc/54AE-MKC7]; Karen Hao, *Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes*, MIT Tech. Rev. (June 6, 2019), https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes [https://perma.cc/CX6R-5C5P].

[18] Arlind Kadra et al., *Regularization is All You Need: Simple Neural Nets Can Excel on Tabular Data* (2021), https://arxiv.org/pdf/2106.11189.pdf [https://perma.cc/59WH-8BX2] (preprint).

[19] Bender et al., *supra* note 15.

[20] *See* Eric Null, Iseuda Oribhabor, & Willmary Escoto, *Data Minimization: Key to Protecting Privacy and Reducing Harm*, Access Now (2021), https://www.accessnow.org/cms/assets/uploads/2021/05/Data-Minimization-Report.pdf [https://perma.cc/99GU-J25C]; Phil Jones, *Refugees Help Power Machine Learning Advances at Microsoft, Facebook, and Amazon*, Rest of World (Sept. 22, 2021), https://restofworld.org/2021/refugees-machine-learning-big-tech [https://perma.cc/3MKF-J46L].

[21] *See* Karen Hao, *Inside the Fight to Reclaim AI from Big Tech's Control*, MIT Tech. Rev. (June 14, 2021), https://www.technologyreview.com/2021/06/14/1026148/ai-big-tech-timnit-gebru-paper-ethics [https://perma.cc/X4ZG-RZ4B].

[22] *See* Jon Harper, *Federal AI Spending to Top $6 Billion*, Nat'l Def. (Feb. 10, 2021), https://www.nationaldefensemagazine.org/articles/2021/2/10/federal-ai-spending-to-top-$6-billion [https://perma.cc/Q24J-3H8D].

[23] *See* Gary Marcus, *Deep Learning: A Critical Appraisal* (2018), https://arxiv.org/pdf/1801.00631.pdf [https://perma.cc/Q5D8-WR86]; François Chollet (@fchollet), Twitter (Dec. 18, 2017), https://twitter.com/fchollet/status/942733414788190209 [https://perma.cc/6E65-VJGN] ("For most problems where

For the same reasons, we believe the establishment of the NAIRR should focus on offering an alternative to the data- and compute-hungry applications that are the focus of many industry and research labs—most of whom already have substantial resources at their disposal and need no additional government support. Instead, the Task Force should emphasize shared data access to researchers, journalists, and advocates who are frequently shut out from access to datasets held by companies or public sector entities. The scarce resource for underrepresented researchers is not technology infrastructure—it is access to relevant, real-world data that companies and public agencies guard closely. Even for those focused on the use of deep learning, there is a longstanding problem with the data sets available to researchers,[24] especially those who are not associated with large technology companies and who may be able to shed light into the risks of this technology or develop new innovations.

We will spend the remainder of this paper focusing on the NAIRR as a shared data repository. We ask that the Task Force especially consider data that would most help civil rights and society groups to inspect and understand widely used AI, the privacy and ethical implications of data collection and/or acquisition, and the risks associated with the NAIRR.

IV. Shared Data that Would Help Advance Civil Rights Concerns and the Ability to Inspect

To advance the protection of civil rights, any shared data repository should contain a number of mandatory datasets as well as accompanying requirements for these datasets. We recommend that the NAIRR Task Force require the inclusion of datasets used by local, state, or federal agencies that have elected to build predictive models/automated decision systems.

Predictive models by these government actors have a wide-ranging impact on individual lives. Examples of predictive models currently being used by local government agencies include pretrial risk assessment tools and child welfare screening tools. Pretrial risk assessment tools are often built on datasets that include demographic information and information about an individual's criminal legal system history.[25] Child welfare screening tools can draw from datasets that include demographic information and child welfare, jail, juvenile probation, behavioral health, and birth records.[26] Journalists and civil rights groups have already raised equity concerns about both.[27]

---

deep learning has enabled transformationally better solutions (vision, speech), we've entered diminishing returns territory in 2016-2017.").

[24] *See* Amandalynne Paullada et al., *Data and Its (Dis)contents: A Survey of Dataset Development and Use in Machine Learning Research*, NeurIPS 2020 Workshop, https://arxiv.org/pdf/2012.05345.pdf [https://perma.cc/7B6G-8CLW].

[25] Mapping Pretrial Injustice, *Inputs: Variables*, https://pretrialrisk.com/the-basics/pretrial-risk-assessment-instruments-prai/inputs-variables [https://perma.cc/F58Z-5YLB].

[26] *See* Oregon Dep't of Hum. Servs., Oregon DHS Safety at Screening Tool–Development and Execution at 4 (2019), https://www.oregon.gov/DHS/ORRAI/Documents/Safety%20at%20Screening%20Tool%20Development%20and%20Execution%20Report.pdf [https://perma.cc/X4ZT-YR56]; Rhema Vaithianathan et al., Allegheny Family Screening Tool: Methodology, Version 2, at 3 (2019), https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V2-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-7.pdf [https://perma.cc/KLJ9-AHAB].

[27] Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/ESW2-B9YQ]; Anjana Samant et al., *Family Surveillance by Algorithm: The Rapidly Spreading Tools Few Have Heard Of*, ACLU (Sept. 29, 2021), https://www.aclu.org/news/womens-rights/family-surveillance-by-algorithm-the-rapidly-spreading-tools-few-have-heard-of [https://perma.cc/M55Z-F2XL].

One aspirational goal for NAIRR would be to offer resources for community-based journalistic and academic researchers to provide oversight on powerful private sector actors that use AI over exclusive/proprietary data. These institutions often shut out marginalized or underrepresented groups looking to understand their impact. A recent example is Facebook's decision to block "good-faith research in the public interest" and shut down the accounts of New York University researchers who were studying how the platform's political advertisers target different users.[28] For these reasons, the government should encourage private actors to offer datasets that would help researchers evaluate the fairness risks of tools that rely on these datasets. The NAIRR Task Force should consider how to provide incentives and support to independent researchers who are engaged in the oversight of privatized AI/ML systems that have significant social impact.

To ensure the usefulness of datasets, the NAIRR Task Force should set standards for data documentation and require civil rights assessments of both the use case and data provenance, as described further in the sections below.[29] To facilitate audits, the datasets should include protected class data such as race and gender when legally permitted, and when it is not, sufficient data to infer protected class status using standardized methods.[30] Contributors of datasets should be required to document inherent flaws of their data (e.g., selection bias) and discuss the possible creation of datasets that overcome these flaws. Datasets should also include qualitative data, including feedback from listening sessions with impacted communities and from tool users such as agency workers.[31] Finally, if the NAIRR Task Force determines that the submitted data or methods are too flawed, it should be prepared to ask an entity to halt or refrain from deploying the predictive model built on that data and should issue an explanatory statement recommending the technology not be used. We hope the creation of NAIRR, if open to a broader set of researchers, can encourage more impact evaluations like the National Institute of Standards and Technology's Face Recognition Vendor Test report or the research reports resulting from the Criminal Justice Administrative Records System.[32]

---

[28] *See* Letter from Samuel Levine, Acting Dir. of the Bureau of Consumer Prot., to Facebook (Aug. 5, 2021), https://www.ftc.gov/news-events/blogs/consumer-blog/2021/08/letter-acting-director-bureau-consumer-protection-samuel [https://perma.cc/68PD-HA4N]; Ethan Zuckerman, *Facebook Cares About Privacy—But Only If You're an Advertiser*, The Atlantic (Aug. 6, 2021), https://www.theatlantic.com/technology/archive/2021/08/facebook-only-cares-about-privacy-advertisers/619691/ [https://perma.cc/6RWH-YBY4].

[29] *See* Timnit Gebru et al., *Datasheets for Datasets* (Mar. 19, 2020) (preprint), https://arxiv.org/abs/1803.09010v7 [https://perma.cc/NL72-R75K]; *The Dataset Nutrition Label*, Data Nutrition Project, https://datanutrition.org/labels/ [https://perma.cc/GD9C-LHXQ]; Bao et al., *supra* note 3.

[30] Regulation B of the Equal Credit Opportunity Act, for example, does not allow the use or collection of protected characteristics, such as race or gender, by financial institutions except for mortgage decisions. *See, e.g.*, Equal Credit Opportunity Act (Regulation B) Ethnicity and Race Information Collection, 12 C.F.R. § 1002 (2017), https://www.federalregister.gov/d/2017-20417 [https://perma.cc/VR8C-D2AH]. The standard approach in the credit and lending industry is to use publicly available information as a proxy for unidentified race and ethnicity. Consumer Fin. Prot. Bureau, *Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity* (2014), https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf [https://perma.cc/HQ2E-PV63].

[31] *See* Kelley Fong, *Getting Eyes in the Home: Child Protective Services Investigations and State Surveillance of Family Life*, 85 Am. Socio. Rev. 610 (2020), https://journals.sagepub.com/doi/10.1177/0003122420938460 [https://perma.cc/P5V7-DJDA]; Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, Wired (Jan. 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality/ [https://perma.cc/786P-JAAM].

[32] Patrick Grother, Mei Ngan & Kayee Hanaoka, Nat'l Inst. of Standards & Tech., NISTIR 8280, Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects (Dec. 2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf [https://perma.cc/PTQ8-UNVE]; Research Reports,

V. Civil Rights Impact Assessment Requirements for Data Receipt and Use

We encourage the NAIRR Task Force to ensure that those who receive data from the program are held to a high standard. This may include the use of civil rights impact assessments and other mechanisms of assurance in any implementation of the NAIRR before providing data, but also by those who intend to use the NAIRR's datasets or research with it. These mechanisms must meaningfully include input from impacted people and communities in their design and execution.[33] As we have written elsewhere, we believe such impact assessments should be conducted according to standards that set out necessary evaluation points, and at a minimum should require: regular evaluation for discriminatory effects throughout the conception and development of any model based on the data, and—if not terminated during development due to unacceptable impacts or for other reasons—in its implementation and use; proactive searches for and adoption of less discriminatory alternatives; continuing assessments of whether the data used in training technologies is representative and accurate; and that the technologies measure lawful and meaningful attributes and seek to predict valid target outcomes.[34]

While some entities are voluntarily evaluating their AI systems, there is too little transparency in the documentation and publication of the resulting impact assessments. The Task Force could help solve this issue by enhancing researcher access to previously unevaluated AI datasets, and it should also work to ensure that any uses of the data come with documentation, transparency, and accountability requirements. When an impact assessment is conducted, it is important that the Task Force require information about the evaluation be made publicly available, including information about the content and reasoning behind the evaluation, who is conducting the evaluation, and what their relationship is to the entity being evaluated, if the evaluation is not conducted by a government agency itself.

If educational materials will be provided to those who use or access the NAIRR, those materials should go beyond mechanical and technical concepts; they should also address issues in equity, AI ethics, and the potential disparate impact of AI.[35] Educational resources should also include details on how to conduct audits for fairness and disparate impact of AI tools.

Crim. Just. Admin. Rec. Sys., https://cjars.isr.umich.edu/overview/research/research-reports/ [https://perma.cc/Z2HD-QCSW].

[33] *See* Dillon Reisman et al., AI Now Inst., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* 18–20 (Apr. 18, 2018), https://ainowinstitute.org/aiareport2018.pdf [https://perma.cc/JQY8-VB4L] ("[Entities] should ensure that affected communities are able to suggest researchers that they feel represent their interests, and should work with researchers to ensure that these communities have a voice in formulating the questions that are asked and addressed by research and auditing.").

[34] *See ACLU Comment on NIST's Proposal for Managing Bias in AI*, ACLU (Sept. 10, 2021), https://www.aclu.org/letter/aclu-comment-nists-proposal-managing-bias-ai [https://perma.cc/B58E-VXF5].

[35] Maria Kasinidou et al., *Educating Computer Science Students About Algorithmic Fairness, Accountability, Transparency and Ethics*, 1 Proc. of the 26th ACM Conf. on Innovation & Tech. in Comp. Sci. Educ. 484 (2021), https://dl.acm.org/doi/abs/10.1145/3430665.3456311 [https://perma.cc/H53F-KBR2]; Veronika Bogina et al., *Educating Software and AI Stakeholders About Algorithmic Fairness, Accountability, Transparency, and Ethics*, Int'l J. of Artificial Intel. in Educ. 1 (2021), https://link.springer.com/article/10.1007/s40593-021-00248-0 [https://perma.cc/RZ5U-3A9K]; Anna Lauren Hoffman & Katherine Alejandra Cross, *Teaching Data Ethics: Foundations and Possibilities from Engineering and Computer Science Ethics Education*, https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/46921/TeachingDataEthicsFoundations-Hoffmann-Cross.pdf [https://perma.cc/HH92-7YPF].

VI. Privacy Implications of Data Collection and/or Acquisition

The NAIRR aims to make creating AI tools more accessible to researchers and other AI practitioners, and this may include making data available for models to be trained and tested. This raises questions of what kinds of data the Task Force believes should be made available. The Task Force will need to make clear the kinds of data it believes are appropriate to be made widely available and what AI tools are appropriate to be developed based on this data. For example, the Task Force will need to decide whether images of faces will be among data it makes available. While numerous studies have detailed that current facial recognition technology is in need of improvement in accuracy, some organizations, including the ACLU, are opposed to the creation of these tools because of their potential use for pervasive surveillance even after technological biases have been addressed. Similarly, the inclusion of public data on arrests or crime rates may be controversial because of the known history of racial and ethnic bias within this data. The Task Force must take care to ensure that data selected to be included in this database cannot be used to perpetuate harm and increase surveillance.

VII. Datasets Present Risks

Distributing access to any dataset may present risks to individuals whose data are present, or to communities of people similar to those present in the dataset. The Task Force should consider these risks to data subjects and have clear plans for how to mitigate them as it looks to expands access to datasets. Examples of risks include identity theft, harassment, discrimination, unwarranted legal scrutiny, surveillance, and financial harm data harms.[36] Such harms can arise from release of medical information, location history, personal imagery, financial records, employment details, sexual history, religious information, and more. Any large, socially relevant dataset can contain sensitive information that can cause harm: either to people or communities represented in the datasets, or to people or communities subject to systems trained on the dataset. The NAIRR will likely collect, hold, and redistribute datasets, and thus has responsibilities as a data steward.

VIII. Redress for Harm and Responsibility for Researchers

It can be difficult for an individual or group to know that a harm they have experienced is due to a dataset released for research. It is also unclear how to identify parties who have been harmed, or how to meaningfully redress harms when they are identified. When a researcher causes harm through misuse of a dataset, inadvertently or on purpose, it is not clear what consequences or liability they face, or how it would be enforced. The Task Force should address these concerns and develop policies and practices to minimize harm, provide redress when it occurs, and structure liability for researchers.

IX. Examples of Attempts at Responsible Data Management

Managers of other large, research-oriented datasets have tried to tackle similarly significant challenges related to privacy and other harms. An example is the National Center for Health Statistics Research Data Center's documentation of anti-disclosure rules for researchers,[37] which include both institutional and individual consequences, and association with legal liability regimes

---

[36] Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, GWU Legal Stud. Rsch. Paper No. 2021-11 (Feb. 9, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3782222 [https://perma.cc/N2H3-6KYH].

[37] Nat'l Ctr. for Health Stat. Rsch. Data Ctr., Preventing Disclosure: Rules for Researchers (2020), https://www.cdc.gov/rdc/data/b4/Disclosure-Manual-v2.5.pdf [https://perma.cc/WEM3-3SVW].

like the Confidential Information Protection and Statistical Efficiency Act.[38] Another example is the U.S. Census Bureau's responsibility for distributing large amounts of analyzable data with potentially serious side effects for the people involved. Recent attempts at safer data distribution policies include differential privacy and other statistical safeguards.[39] These privacy preserving techniques may also reduce the accuracy of the data or may limit its applicability for certain kinds of use. In addition, Duke University has published a dataset known as DukeMTMC aimed at improving vision analysis, person-recognition, and object tracking. Duke later took down that dataset[40] in response to a report about its widespread use[41] and associated harms, but the data continues to be used despite the retraction.[42] While these examples include attempts to mitigate harm, the efficacy of these mitigations is unclear, as is researcher liability when datasets are misused. The Task Force should ensure that the NAIRR does better at protecting data subjects, redressing harms, and ensuring consequences for researchers whose work causes harm.

We thank you for considering our suggestions.

Sincerely,

The American Civil Liberties Union

---

[38] Confidential Information Protection and Statistical Efficiency Act of 2002, 44 U.S.C. §§ 3501-3521, https://www.eia.gov/cipsea/cipsea.pdf [https://perma.cc/G83Z-JZQQ].

[39] Simson L. Garfinkel et al., U.S. Census Bureau, Differential Privacy at the US Census Bureau: Status Report (Jan. 27, 2020), https://csrc.nist.gov/CSRC/media/Projects/pec/documents/stppa-01-20200127-talk03-Garfinkel-diff-priv-census.pdf [https://perma.cc/AG7K-JX5R]; U.S. Census Bureau, Statistical Safeguards, https://www.census.gov/about/policies/privacy/statistical_safeguards.html [https://perma.cc/TA4P-QHCP] (last visited Sept. 30, 2021).

[40] Jake Satisky, *A Duke study recorded thousands of students' faces. Now they're being used all over the world*, Duke Chron. (June 11, 2019), https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur [https://perma.cc/X2L8-V54L].

[41] *Duke MTMC Dataset*, Exposing.ai, https://exposing.ai/duke_mtmc/ [https://perma.cc/VEP7-D7J8] (last visited Sept. 30, 2021).

[42] Kenny Peng, *Facial recognition datasets are being widely used despite being taken down due to ethical concerns. Here's how.*, Freedom to Tinker (Oct. 21, 2020), https://freedom-to-tinker.com/2020/10/21/facial-recognition-datasets-are-being-widely-used-despite-being-taken-down-due-to-ethical-concerns-heres-how/ [https://perma.cc/FD93-LNQU].