Social and Ethical Implications of Human-Centered Artificial Intelligence (AI)

Date: 06/02/2019

Re: NIST RFI: Developing a Federal AI Standards Engagement Plan

Submitted by: Rhonda J. Moore, PhD, MA US DHHS-FDA, Rhonda.Moore@fda.hhs.gov

Disclaimer: This document is not a formal dissemination of information by the US Department of Health and Human Services (US-DHHS), and the US-DHHS FDA and does not represent agency position or policy.

Most organizations do not yet have a comprehensive framework for evaluating the ethical and social dimensions of Human-Centered AI, or a process for understanding the context-specific impacts of emerging AI on regulatory research, review and decision making. The general social and ethical considerations outlined below are expected to enhance federal benefit utilizing AI standards.

Comprehensive Code of Human-Centered AI Ethics: Human-Centered AI should be developed for the common good and benefit of humanity. All people of diverse backgrounds should be able to flourish mentally, socially, emotionally and economically alongside human-centered AI. However, with advances in technology; technical skills may erode, become increasingly become obsolete; perpetuating existing and enduring inequalities, particularly in those most vulnerable. A code of human-centered AI ethics is a set of agreed upon values, a framework and a guiding light for action. In the context of social and ethical implications of Human Centered AI, this code should not be restricted to only data collection and data governance efforts. Rather it includes: what the data is, how data is collected, who are the creators/ designers and users of the AI, how the data are interpreted and acted upon. It exists as part of the broader framework for existing and future innovation efforts (e.g., lifecycle of innovations, product design decision making, and regulatory review, hiring, training and retention of a diverse and agile workforce), infrastructure developments (e.g. cybersecurity, IT), governance strategies (e.g. data management strategies that are robust, known and understood by diverse team members, reviewed regularly, adapting as a company grows and changes), also integrated into training opportunities for diverse staff, and across all human resource staffing efforts.

Tenets of Ethical Human Centered AI: Role of Cultural Values: Values such as fairness, transparency, accountability, reliability, openness, trustworthiness, responsibility, and intelligibility should be fully incorporated as part of the culture of an organization and incorporated as part of the framework into in the design, development and implementation of human-centered AI regulatory research, review and decision-making processes, procedures, and in communications with the public.

A code of Human-Centered AI ethics and the tenets of ethical AI should be embedded in into existing and future AI regulatory efforts, which can also be applied toward the following RFI topics of interest to NIST as follows:

Areas of potential interest:

1) Data
    - AI Ethics including addressing bias in data requires a focus and sensitivity to the contexts, methods and instruments of data collection. However, there should be

agreement on which data are relevant to a given problem and decision. It is also important to determine how such judgements about what data are relevant to include, by whom, under what conditions, determining also quality standards, and which data are not included –as these questions will all impact the ability to generalize research findings, and limit regulatory review decision making processes. For instance, a lack of diversity in training data (e.g. lack of datasets labelled by ethnicity) will lead to inequitable accuracy in classifying individuals/groups in different categories potentially impacting regulatory research and review outcomes.

2) Analysis of data
    - Data scientists spend most of their time choosing, gathering, combining, structuring, and organizing data to enable a specific AI to generate meaningful patterns. While reviewers who will be utilizing AI tools in their analysis will benefit from the data meeting high quality standards; having policies and procedures that ensure fairness, transparency, accountability, reliability and reproducibility is also central to the data curation, and the analysis of data as part of regulatory review and regulatory decision making.
3) Data sources
    - The utilization of reliable data sources increases the speed, retrieval and delivery of information, which can also potentially increase trust and reduce the overhead of data quality validation. However, significant barriers persist, and it is important to include the perspectives of diverse decision makers and stakeholders to agree upon what is *appropriate* source data whilst also ensuring the privacy of research participants and safety of data. This also includes protecting individuals from harm based on algorithmic or data bias or unintended correlation of personally identifiable information (PII) even when using anonymous data.
    - In addition, while standard methodologies for identifying, classifying and ranking "appropriate" curated data sources of potential value to that were obtained or processed with AI-assisted technologies is important; alternative methodologies (e.g., piloting new and emerging techniques) should also utilized to also contribute to the evaluation and comparison of best practices of fairness and governance models to effectively prevent, identify, and address bias. These strategies will also substantially improve appropriateness, accuracy, transparency, fairness and efficiency in data sourcing.
4) Methodologies
    - Any results of AI products including ML algorithms in all stages of submission should be produced using industry-standard methodologies to ensure interpretability and replicability. These industry standards are currently being developed and many are often context specific. As an initial step and perhaps in tandem with organizational recommendations for regulatory science studies, it is important to also ensure that AI designers are also aware of potential bias in the development of their algorithms and include individuals from diverse and underrepresented populations in their user testing.
5) Intelligible AI & Algorithmic Literacy

- Data privacy is important, and consumers are conscious of the sharing of private data. Challenges of data privacy and security impact how an organization approaches data collection and sharing. Along with new global regulatory efforts (e.g., GDPR, Article 22), companies are also having conversations around these efforts including AI ethics and privacy.
- Transparency is critical to enable the real-world deployment of intelligent systems including ML to ensure that ML models are working as intended. Also, while there is broad awareness of AI, there are also inaccurate understandings across diverse segments, as to what human-centered AI is, what is does and diverse impacts on privacy, collection and security of data. As part of a comprehensive AI ethics plan; the data management strategy of an organization should be robust, known to all diverse team members, reviewed regularly, adapting as the organization grows and changes.
- Engagement and enhancing user trust across diverse stakeholders, including citizens, governmental bodies, academia (e.g. STEAM fields), industry, and across diverse public and private entities should also be based on accountability, transparency, trustworthiness, openness and fairness.

6) Inclusive AI: Human- Centered AI Design, Training and Human Resource Management
   - AI is currently being used to accelerate and assist inclusive human-centered AI efforts in design, training and human resource management; across a variety of high-volume tasks such as: narrowing the talent pipeline (e.g. improving candidate vetting), automating business processes, and replacing repetitive administrative tasks. Barriers to successful implication exist including the lack of inclusion of diverse decision makers. Continuing to utilize interdisciplinary and diverse subject matter expertise to identify, analyze and apply AI-assisted technologies is critical to the development of inclusive human-centered AI integrated into design, training and human resource efforts (e.g. addressing fairness and bias in candidate resumes).
   - The development of targeted AI interventions to identify and address the risk of unintended consequences and context effects (e.g. historical bias and exclusion criteria that continue to impact the analyses of data and decision making) is also critical in the development of inclusive and ethical AI.
   - The Human-Centered AI conversation is clearly much bigger than the fields of computer science and information technology. Human-Centered Design, Training and Human Resource Management in AI should be interdisciplinary and include diverse subject matter expertise and diverse representation of individuals from STEAM fields (e.g. humanities and social sciences). There are also several on-going academic and industry led efforts that also shed light on these issues in the states and overseas (e.g. Stanford Human Centered AI Institute, MIT Human Centered AI, IBM AI Fairness 360 Toolkit, AI Now Algorithmic Policy Toolkit, etc.).

7) Next steps:
   - The government should continue to convene interdisciplinary meetings and summits of diverse stakeholders including governments (e.g. state, local, federal, international), academia (e.g. STEAM fields), citizens, and industry across current and evolving future

areas as part of a long-term effort to establish norms and standards for the societal and ethical design, development, regulation and deployment of Human-centered AI. These plans should also include strategies to manage, monitor, and mitigate unintended consequences and risks in both the short and long term.

Select References:

1. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York St. Martins Press. 2018.
2. O'Brien, A. IDC PlanScape: Responsible and Ethical AI for Federal and State Governments. 2019. Retrieved from: https://www.idc.com/getdoc.jsp?containerId=US44856318.
3. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books; Reprint edition (September 6, 2016).
4. Lohr S. Facial Recognition Is Accurate, if You're a White Guy. New York Times. February 9, 2018. Retrieved from: https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.
5. Ho CWL, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. Clin Radiol. 2019 May;74(5):329-337.
6. Weber C. Engineering Bias in AI. IEEE Pulse. 2019 Jan-Feb;10(1):15-17.
7. McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics. 2019 Mar;45(3):156-160.
8. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans A Math Phys Eng Sci. 2018 Oct15;376(2133).
9. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature. 2018 Jul;559(7714):324-326.
9. Banks J. The Human Touch: Practical and Ethical Implications of Putting AI and Robotics to Work for Patients. IEEE Pulse. 2018 May-Jun;9(3):15-18.
10. Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL, Fei-Fei L. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. Proc Natl Acad Sci U S A. 2017 Dec 12;114(50):13108-13113.
11. Howard A, Borenstein J. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. Sci Eng Ethics. 2018Oct;24(5):1521-1536.
12. Stylianou N. Ethics must be at centre of AI technology, says Lords report. Retrieved from: https://news.sky.com/story/ethics-must-be-at-centre-of-ai-technology-says-lords-report-11333333. Date accessed: December 15, 2018.
13. West S, Whittaker M, Crawford K. Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. 2019. Retrieved from: https://ainowinstitute.org/discriminatingsystems.html.
14. AI Now. Algorithmic Accountability Policy Toolkit. 2018. Retrieved from: https://ainowinstitute.org/aap-toolkit.html.
15. IBM AI Fairness 360 Toolkit. 2018. Retrieved from: https://github.com/IBM/AIF360.