Comments of the World Privacy Forum to the Secretary's Advisory Committee on Human Research Protections (HHS) regarding the SACHRP AI Framework

Sent via email to: diek@uw.edu, Jerry.Menikoff@hhs.gov, SACHRP@hhs.gov

July 20, 2022

Douglas S. Diekema, M.D., MPH
Chair, Secretary's Advisory Committee on Human Research Protections (SACHRP) Professor, Department of Pediatrics
Adjunct Professor, Departments of Bioethics and Emergency Medicine
University of Washington School of Medicine

Jerry Menikoff, M.D., J.D.
Director, Office for Human Research Protections (OHRP) 1101 Wootton Parkway, Suite 200
Rockville, MD 20852

Re: Comments regarding SACHRP AI Framework

Dear Drs. Diekema and Menikoff, and members of the Committee:

The World Privacy Forum thanks SACHRP for the opportunity to review its AI Framework, and for providing the opportunity to submit written comments. The World Privacy Forum (WPF) is a non-profit public interest research group based in the U.S. We publish significant materials focused on privacy. See www.worldprivacyforum.org. WPF has dual expertise in health privacy and in AI/ML. We have published substantive work in both fields, including peer-reviewed work. In addition to the research we conduct, we are an original member of the OECD's AI Expert Group, and were part of the team that created the OECD Artificial Intelligence Principles. Currently, WPF is the lead for the civil society work on AI at the OECD AI Working Part, including work on accountability in AI structures and systems. Additionally, WPF co-chairs the World Health Organization's Research, Academic, and Technical Network (with the CDC).

Human subjects research that also utilizes Artificial Intelligence / Machine Learning techniques (hereafter AI/ML) is evolving at a rapid pace, and has gone beyond just a few instances of exploratory research. WPF agrees that the Common Rule requires explicit interpretation regarding AI techniques.

For these comments, we proceed section by section of the Framework. We have skipped some portions. Our comments reference the document, *AI Framework*, available via regulations.gov under the SACHRP docket for 20-21 July, 2022.

I. Comments regarding *Defining AI* *(Lines 15-43)*

The definition of AI and Machine Learning in the AI Framework requires additional work. There are well-understood and widely accepted definitions of AI/ML that were crafted after lengthy multinational, multi-stakeholder deliberation. As a result, these definitions are precise, technically accurate, and comprehensive. We urge the Committee to adopt one of these definitions and adapt it as necessary.

The OECD definition of AI/ML was crafted in consultation with leading global experts on AI/ML, and included over 50 experts from government, the technical community, civil society, business, standards development organizations, trade unions, and formal stakeholder groups, including business, civil society, and standards development organizations, among others. This work at the OECD took years, and was highly deliberative and took place under formal multilateral due process rules. The U.S. formally ratified the OECD Guidelines on AI, which are now soft law in the U.S. — and in other ratifying countries — as a result.

Here is the OECD definition of AI/ML systems:

> An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.

We note that the emphasis in the OECD definition is appropriately focused on a systems approach, which is an important component of a thorough analysis of risks and other aspects of AI/ML systems, including those involved in sensitive research.

In line 28, the Framework discusses that Artificial Intelligence can be a misleading term. We agree. However, "AI" has become part of the mainstream parlance. As a result, we support the use of the term "AI/ML" to discuss what is in reality mostly ML systems. We encourage SACHRP to not use the term AI on its own, as it is indeed not a precise use of the term, as the Framework acknowledges.

II. Comments regarding *AI and "Big Data"* *(Lines 44- 69)*

For this Framework, we encourage avoiding the term "big data." Even if 'big data" was not aging rapidly in the vernacular, we would still encourage a more precise use of terms around data used in AI/ML processes.

Beyond this point about excising the term "big data" from the Framework, we have several additional comments regarding this section:

A. The overall explanation of how AI/ML data techniques push against the Common Rule is quite good, and we support it. Again, however, we urge the removal of the term "big data" from this section. ML systems operative in the health research sector do not always rely on high volume or high velocity data. There are instances when demonstrably small but important data sets may be utilized.

B. The discussion of privacy harms in this section is not complete, and as it currently exists, the discussion may provide a misleading presentation of how privacy harms occur in the AI/ML context. Privacy harms arise from multiple factors in AI/ML systems. The harms can come from

too much data, and the harms can come from too little data. The harms can come from historical data that is accurate, and the harms can come from data that is inaccurate, historical or otherwise. Harms can come when data is fully complete. Harms can arise also from incomplete data. Harms can come from algorithms with one or more pathologies, and harms can come from faulty hardware components. Harms can come from improper fit of the algorithm, and harms can come from improper handling of outliers. Harms can come from improper uses or interpretations of the data. There are additional harms, and we urge the committee to more carefully consider what these might be looking at the full AI/ML lifecycle.

By way of providing some concrete exemplars of privacy harms relating to AI/ML, we point you to our report, *The Scoring of America,* which was 7 years in its research. https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/ This report was cited by the White House and the FTC at the time of its publication, and it still stands as a groundbreaking report on precisely how AI/ML can harm privacy of individuals and groups of individuals. There are several sections of the *Scoring* report that focus on data in the health context. We note that broadly speaking, harms relating to classification errors needs to be more prominent in the Framework, and could be fit readily into this section of the Framework that discusses data.

C. The discussion of data in this section of the Framework includes some comments on identifiability. We agree with this statement in the Framework:

> 63 It should be clear that a given data set can be both
> 64 overly broad from the perspective of an individual (a BD issue) and incomplete from the
> 65 perspective of a biological mechanism or an inadequately sampled population.

We agree with the argument in this section that the issue of identifiability takes on a different role in the risk analysis in the context of the Framework. Most records are becoming increasingly identifiable over time for a variety of reasons, some of which are technological in nature. Even those records which are not identified at the time of initial analysis can still be analyzed and can potentially be applied to individuals at the point of end use through a variety of techniques.

III. Comments regarding *AI and software as a medical device* *(Lines 70-98)*

The discussion in the Framework of *in vitro* diagnostic devices is extremely important. (Lines 70-98). We agree with analysis that AI can be used as a clinical equivalent of an IVD, "basing its diagnostic outcomes on data rather than biological specimens." (lines 83, 84). We agree that poorly planned AI studies can impact populations and groups of people. They can *also* impact individual research participants, particularly if the AI/ML is classifying or scoring, or otherwise making decisions about outliers.

Software as a Medical Device (SAMD) will likely prove to be an important regulatory framework for AI/ML, and we agree that will be important to ensure that the Common Rule, with its broader scope than the FDA, is informed by the FDA's approach.

IV. Comments regarding *Potential harms of AI* *(Lines 192-225)*

We reiterate that harms from AI/ML in the context of the Common Rule do not arise only from unrecognized limitations or biases in data sets. There is a complex web of harms that can arise from multiple sources, and at different points in the AI/ML lifecycle. These problems can span the gamut from software to hardware issues, and can also encompass the whole of the AI/ML

lifecycle and each of the relevant AI/ML ecosystems that are interconnected with the system(s) being utilized. The problems can also relate to imported use of pre-fabricated algorithmic blocks that are produced elsewhere and may be opaque to the researchers using them. Additional obstacles can be posed by the method of analysis, the fit of the algorithms or network of algorithms used, the age of the algorithm, and many additional factors. These ideas are well-substantiated in broader AI/ML work.

We again agree that AI/ML harms can impact groups of people. We also reiterate that AI/ML's possible harms are not only inclusive of "intrinsically group harms that arise from biases in the original data set." (Lines 211-212). The harms of AI/ML in the Common Rule context are inclusive of this problem, and also include a variety of individual-level issues. Sometimes individual and group harms can be mixed together in complex ways. We reference the harms experienced in aggregate credit scoring as an important use case to understand, as it illustrates the challenging complexity posed by AI/ML that can be applied at a group and at an individual level.

Aggregate credit scoring is where Census blocks or households are given scores as to their creditworthiness. These scores typically include more than 1,000 factors from a wide variety of data sets. The aggregate credit scores are not created for applicability to any individual, or they would run afoul of the Fair Credit Reporting Act, which regulates credit scoring in multiple ways and applies to *individual* consumers. Because aggregate credit scores apply only to *groups* of people, and not to individuals, these types of credit scores are unregulated. However, in actual use, aggregate credit scores that are designed to apply at the household or Census block level may in practice also be applied to individuals by some end users, especially for marketing purposes.

What is the harm here? The harm is that an algorithmic decision-making process that is regulated under the FCRA both in allowable factors used, and in the algorithms used, evades these regulatory guardrails and can use data containing any number of biases or problems, and can use algorithms that are not subject to regulatory oversight, and can apply them in ways that would not be allowable under the FCRA — and this can all happen because the aggregate scores are specifically designed to apply to groups of people and not individuals. Further, the unregulated aggregate scores that are designed to apply only to groups of people may in actual practice be applied to individuals.

We cannot stress enough that the major harms of AI/ML do not arise only from imperfections in the original data set. Yes, data sets can have problems and they are certainly one important source of harms, and we do not want to diminish this. But data set problems are far from the only causes of potential harms.

For example, the actual quality and performance of the algorithm(s) can display pathologies that are not dependent just on a flawed data set. In NIST testing of industry face recognition algorithms across an exceptionally clean, systematic, and complete data set, the algorithms themselves displayed a variety of problems and pathologies. In a famous NIST study on demographic differentials in face recognition systems, the dataset was removed as the problematic factor, which allowed NIST to study the algorithmic deficiencies with specificity. In its study, NIST processed "18.27 million images of 8.49 million people through 189 mostly commercial algorithms from 99 developers." Even though this NIST study is not focused on human subjects research, a careful reading could be helpful to the Committee, particularly the 1,200 pages of charts that report results for each algorithm. (https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf).

Beyond algorithmic deficiencies, harms can arise from improper or inconsistent interpretation of the outputs, which involves level-setting and a great deal of judgement. In credit scoring, the interpretation and uses of a credit score created under the FCRA are regulated. This is why a credit score 780 means something precise, as does a credit score below 600. This is also why credit scores can only be used in certain specific ways.

What about human subjects research harms? The Framework needs to work through the full roster of possibilities here, from data set issues to algorithm issues to interpretation of AI/ML outputs to uses of AI/ML outputs to handling of outliers in the outputs. There are many factors that can cause harms and problems in AI/ML systems as applied in the human subjects research context. The full scope, range, and consequence of these harms should be articulated in this Framework.

We urge the Committee to bring in experts with multi-disciplinary expertise in key AI/ML systems from other domains in order to learn from existing use cases and see what lessons learned on a technical and policy level can be mapped to human subjects research and then assessed for the potential harms and the protections that will be needed in the domain SACHRP is working to address.

V. Comments regarding *Regulatory areas of concern* (Lines 226-259)

The Framework states that AI/ML raises concerns about three particular tenets of the Common Rule. The emphasis in this section continues to be on **group harms**. We again agree that group harms are a significant issue. We again insist that AI/ML systems in this context can **also include harms to individuals**. The problems that AI/ML bring to groups and to individuals in the human subjects research context must be taken into consideration, because both may be present.

The Framework questions the exemption in 45 CFR 104(d)(2) for certain types of research. We agree. There is a need for protections to be put in place to mitigate harms to groups of people, and / or to people who have been classified in a particular way by the use of AI/ML. It is important to address that AI/ML processes can create new groupings of people by classifying or scoring, and that these new groups or classifications may in and of themselves be problematic. To reiterate, the Framework correctly understands that defined groups of research subjects or people can be harmed. The Framework also needs to encompass the issue that the use of AI/ML in a human subjects research context can also create new groups, which needs to be evaluated for harm. And then finally, there is still risk of harms to individuals. How outliers are handled and understood within the research outputs is an important component to include for harms analysis.

VI. Comments regarding *Framework Recommendations* *(Lines 436-528)*

A. Identifiability and Privacy: We agree and support the idea that AI/ML research exposes the limits of identifiability (or de-identifiability) that form part of the basis of privacy protections under the Common Rule. The addition of a discussion of synthetic data in this section would be useful.

B. Definition of human subjects: It highly problematic that data deemed public - including data that became public due to dissemination by the data subject(s) - is exempted from Common Rule protections. However, as problematic as this was in the past, today this issue has become an untenable policy. Social media disclosures form one aspect of the problem.

There are additional aspects. One issue is that data compilations using GPS data obtained with "click through" consents via apps on mobile phones can now collect detailed enough data to create mappings of individuals' daily lives. This data is collected in ways that may be considered public by some definitions, and is the type of data that could potentially be utilized for a variety of human subjects research. However, there is an urgent need for a more precise definition of public versus private behavior, and a discussion of what data are considered private. Here, we cite our *Scoring of America* report, which discloses a lengthy list of confirmed data sources for the use of AI/ML models. The list is based on empirical data, and is quite substantive. It will give a new appreciation for the challenges of divining a line segregating public and private data. This is a complex, challenging topic which deserves a front-of-the-line spot in SACHRP discussions.

We observe overall that that "consent" is not as simple as it used to be. Individuals confronted with requests for consent from dozens of apps and hundreds (or thousands) of websites in the course of a month or a year rarely have any idea what they consented to. Consents for human subjects research have the same problem of comprehension, voluntariness, and consent fatigue. Polls show that many consider protecting their privacy is hopeless, and they are likely to consent to anything on the grounds that all hope is lost anyway.

C. The necessity of inclusion in setting new standards: We support the need for inclusion discussed in the Framework. We urge that this inclusion extend to all vulnerable populations and groups, and we also urge that this inclusion extend to people who have been newly categorized, classified, or grouped as a result of AI/ML activities. In the human subjects research context, the impacts will differ based on the context of the research conducted, among other factors. It will be important to develop some exemplar use case models regarding inclusion that researchers in the human subjects research context can utilize as they think through potential risks of exclusion in their work.

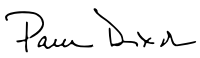
VII. WPF Recommendations to SACHRP

1. We urge that as SACHRP works to create formal guidance regarding AI/ML in the human subjects research context, that harms to individuals and harms to groups of people are studied, and that empirical data from numerous case studies is used to form the basis of recommendations. We hope that this activity would encompass many fora, and take the appropriate time needed to get robust and inclusive feedback from diverse and relevant stakeholders.
2. Many research studies already use AI/ML, which in other contexts would be considered human subjects research. We urge that the SACHRP confront the various loopholes allowing human subjects research to occur without privacy, fairness, meaningful consent, and without protections for other ethical failures. This is a matter of some urgency, as the capacity for such research has outpaced the policy guiding it. Case studies regarding these practices would be helpful in understanding the needs and context for broader guardrails in a changed research environment.
3. We urge SACHRP to be inclusive in its understanding of data sets and their uses. We note that rare and orphan disease experts have a particular interest in AI/ML work, which brings us to the set of challenges regarding sparse datasets. Sparse datasets present different problems in the AI/ML context than high-dimension data sets do, and there has been progress in addressing the problems of sparse datasets. It is important that the Framework address the full range of data sets that may be utilized in human subjects research, and does not focus only on large. high-dimension datasets.
4. The SACHRP has a unique position and can make nuanced recommendations as human subjects research experts. The field of AI/ML is extremely complex, however, and is a discipline in its own right. We encourage SACHRP to take advantage of the expertise in AI/

ML as it applies to large and small populations, and in complex and interconnected ecosystems in multiple use cases. We request that the SACHRP works to create a curated set of case studies specific to multiple aspects of human subjects research that utilizes AI/ML.

Working to address the complex issues that AI/ML systems raise in the context of human subjects research is among the most important of our time for anyone who cares about the balances and integrity needed for human subjects research. Arriving at the proper balance will take time, and will best be achieved with many stakeholders working together cooperatively to understand how integrity and balance can be achieved in the face of profound technical advancements and challenges. Thank you for your work to address these complex and important issues.

Sincerely,



Pam Dixon
Founder and Executive Director,
World Privacy Forum