Open Technology Institute
New America
1899 L Street NW, Suite 400
Washington, DC 20036

OPEN TECHNOLOGY INSTITUTE

August 15, 2014

Tiffany George
Federal Trade Commission
600 Pennsylvania Avenue, NW
Washington, DC 20580

Re: Project No. P145406, *Big Data: A Tool for Inclusion or Exclusion?*

Dear Ms. George,

As the Federal Trade Commission (FTC) formulates the final agenda and speaker line-up for its workshop "Big Data: A Tool for Inclusion or Exclusion?", New America's Open Technology Institute, Data & Society Research Institute, and Solon Barocas, an independent researcher, would like to briefly present a few key concerns and perspectives we think should be included in the workshop's discussions:

1. **The difficulties of gathering empirical evidence demonstrating harm to members of vulnerable populations.**

   When presenting a collection of studies, news reports, and stories about the particular harms from big data that affect diverse and low-income populations (referred to here as "marginalized individuals"), the FTC workshop should address the ethical, social, technical, and legal issues that can hamper the discovery or quantifying of such harms. These difficulties may include, but are not limited to:

   • the ethical problems raised by waiting until harm has occurred to the marginalized individual;

   • the way in which embarrassment or shame (e.g., having been "taken for a ride" or a victim of predatory targeting) contributes to the underreporting of harms by marginalized individuals;

- the challenges that privacy laws, intellectual property laws, computer crime laws, or other potential legal restrictions may present to researchers who could perform large-scale experiments on commercial platforms to detect harm;

- the technical challenge of identifying harms that spring from non-obvious problems with the analysis of big data, as, for example, in cases with unrecognized biased data collection; and,

- the interaction between waiting for more evidence of harm and deepening dependencies upon big data systems (including potentially flawed ones) in the social and economic life of the poor and disadvantaged.

Addressing the limitations of empirical evidence will allow the FTC workshop to consider alternatives in assessing the harms and benefits of big data systems to diverse and low-income populations.

2. **The potential for "bad" big data outcomes despite a lack of "bad" intentions.**

When discussing the wide range of uses of big data systems in the social and economic life of diverse and low-income populations, the FTC workshop should address the role of intentions—of developers, owners, or users of big data systems—and the relationship between intentions and the effects of big data analysis on members of diverse and low-income populations. Specifically, the FTC workshop should address instances where, for example, data scientists or product developers do not intend for big data algorithms to malign specific groups, such as protected classes, but nevertheless lead to outcomes with negative consequences. The workshop should examine both the technical and legal aspects of these kinds of unintentional harm or discrimination, particularly those that stem from the way algorithms 'learn' from data of frequently inconsistent quality and limited scope. A greater focus on actual outcomes will help the FTC workshop explore when and how intentions should matter in the determination of harm, including whether different kinds of intentions (or lack thereof) should imply different kinds of liability.

3. **<u>The unique context of vulnerable populations and the implications for problem solving and next steps.</u>**

When considering the legal landscape and mapping next steps, the FTC workshop should address the unique challenges faced by vulnerable populations in availing themselves of mechanisms designed to minimize harmful outcomes of big data systems. These challenges prevent members of vulnerable populations from taking advantage of remedies available to them. For example, mitigation solutions that require digital literacy will not be helpful to vulnerable populations who lack basic digital literacy skills. That includes downloading a piece of software or consulting an online resource to learn about the potential harms of data profiling. To the extent possible, the FTC workshop should also explore whether attempts to address such barriers such as low digital literacy can result in more effective means of preventing harmful outcomes for all consumers.

This workshop represents a formative moment in ascertaining the breadth and depth of challenges and opportunities of services and products that use big data techniques. It promises to provide further detail and clarity on the issues regarding harm and discrimination in the recent White House report, "Big Data: Seizing Opportunities, Preserving Values." We look forward to participating in the conversation.

Respectfully submitted,

Seeta Peña Gangadharan
Senior Research Fellow
Open Technology Institute
New America
gangadharan@opentechinstitute.org

danah boyd
President and Founder
Data & Society Research Institute
danah@datasociety.net

Solon Barocas
Post-Doctoral Fellow
Center for Information Technology Policy
Princeton University
sbarocas@princeton.edu