A Social Justice Framework for Leveraging Data Science to Advance Gender Equity

Lauren Klein, Director, Digital Humanities Lab, Emory University and Brandeis Marshall, Founder and CEO, DataedX Group

Summary of text:

In today's world, data is an instrument of power, with myriad applications ranging from medical research to response efforts in humanitarian crises. While corporations, governments, and other well-resourced institutions possess the ability to design and deploy data systems, those whose lives and livelihoods are most dependent on the output of these systems remain largely absent from the conversation. This unequal balance of data power can result in interrelated and intersectional harms, including discrimination and increased surveillance. It is possible to leverage data science to advance gender equity, but only if data science and research are ethically and intentionally envisioned from the start.

The unequal balance of power with respect to data science consistently emerges in decisions about what kind of data is collected, what research is undertaken based on that data and how the data is categorized. The interests of those with power, including corporations and governments, disproportionately influence the range of issues addressed via data science. For example, in the United States, women were wholly excluded from medical trials until 1993, due largely to gender bias, as well as concerns over fertility, reproduction and women's fluctuating hormonal levels. This has resulted in generations of medical research which reflects no meaningful sex differences in terms of prevalence of illness, response to treatment and severity of outcomes.

Key recommendations:

- The social, political and historical context surrounding any dataset must be acknowledged through documentation and other qualitative forms of information gathering. Attending to the context of any particular dataset leads not only to more accurate and more truthful data analysis, but also helps to ensure the efficacy and appropriateness of any intervention developed in response to that analysis. Similarly, the ways in which structural power impacts the creation of datasets and data systems must be acknowledged and accounted for.

- Impacted community members must be included as co-designers in any data science project. Participatory design processes such as these can help to ensure data-scientific research is directed towards the issues and opportunities desired by the communities themselves.

- Transparency should be codified through meaningful audits, impact assessment, and individual and collective reflexivity, with the goal of making clear the outcomes and impact of datasets to prevent digital harms. Institutions must be held accountable for the failures and harms of data systems through forceful legal, financial and technical consequences.

- The different forms of labor involved in data work must be credited and compensated. Many forms of data work, on which gender equity increasingly depends, take a psychological toll on researchers.