



Report capturing observations and recommendations from the November 2023 Workshop at UCSF

Report prepared by Workshop Steering Committee:

Paula Fleisher, MA
Gretchen Kiser, PhD
Tung Nguyen, MD
Angela Rizk-Jackson, PhD
Benjamin Rubin, PhD
Ida Sim, MD, PhD
Keith Yamamoto, PhD

Introduction

On November 7, 2023, UCSF hosted a workshop, *Toward Algorithmic Justice in Precision Medicine*, with an overall goal to improve health outcomes and health equity for the steeply increasing proportion of UCSF patients whose care is impacted by algorithms. The workshop had three specific aims: to raise awareness about the scope and impact of algorithmic bias in healthcare and medicine; to learn from experts and invested communities about priorities, concerns, and prospects for solutions; and to help UCSF organizations determine relevant follow-up actions.

“The tech world talks about the technology stack. I want to talk about the value stack. Values all the way down” – Workshop Panelist

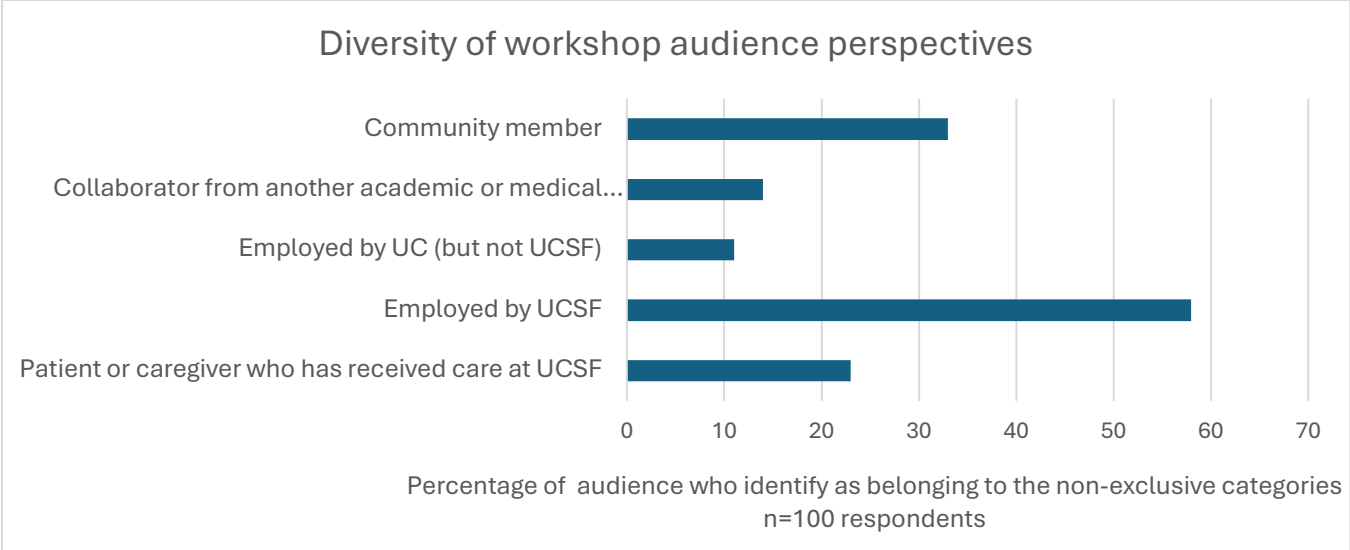
The workshop was organized during a time of rapid advancement in artificial intelligence (AI) technology. The creation, use and impact of algorithms, including those reliant on AI, are expanding dramatically in biomedical research, public health, and healthcare, with potential effects – positive and negative – largely undetermined. As technologies and their applications continue to evolve, movement toward algorithmic justice – assurance of fairness, equity, and accountability in the development and deployment of algorithms and AI systems – will require revision and/or replacement of existing policies and practices, as well as generation of new ones, including effective mechanisms for oversight and accountability.

Three themes, a subset drawn from proceedings of a March 2023 meeting, *Impacts of Healthcare Algorithms on Racial and Ethnic Healthcare Disparities*, hosted by The Agency for Healthcare

Research and Quality (AHRQ) and the National Institute on Minority Health and Health Disparities (NIMHD), framed our workshop:

- 1) Advance toward transparency and explainability in healthcare algorithms and their use.
- 2) Engage patients and communities in all phases of the healthcare algorithm lifecycle and earn trustworthiness.
- 3) Ensure accountability, equity, and justice in outcomes from healthcare algorithms.

The design and implementation of the UCSF workshop reflected an imperative to engage a broad set of community members with vested interest in this topic, including those who are subject to, use, or develop algorithms in research, clinical care, and population health management, as well as those who fund, regulate, or communicate about AI and other algorithms in precision medicine (further referred to in this report as “invested communities”). Two features of the workshop distinguished it from most related efforts: the overarching justice framework and the engagement of community members and patient advocates.



Audience participation tools were employed to capture questions, comments and sentiment regarding certain specified topics. These tools included the virtual meeting application (Zoom Q&A, Zoom transcript), as well as Slido, a web-based audience interaction software (multiple choice and rank polling, word clouds, free text comments). Additionally, during the event’s Breakout Session, notes were generated by each group and later organized and digitized. A full capture of this information can be requested; however, we have taken some of the most common themes and salient points from this output and formulated them into the observations and recommendations found in this report.

The presentation of the workshop recommendations are:

- Categorized and distilled into the [framework topics adapted from the Berdahl et al. comprehensive review paper](#);
- Mapped back to workshop themes;
- Presented with the understanding that details for implementation will fall to subsequent governance body/bodies; and
- Associated with a set of potential UCSF influencers and implementors.

What if we're not achieving equity in the same way that we would ideally like to achieve equity, but it's still better than standard of care, where we are improving people's lives? Those are difficult questions.
– Workshop Panelist

A consensus, universal recommendation was that just, equitable, and ethical use of algorithms and analysis tools in health research and healthcare will require effective development and communication of best practices and standards in order to support a shared understanding of the many issues and trade-offs.

Important Note: This report is not intended to comprehensively capture all the issues relating to Algorithmic Justice in Precision Medicine, rather, it documents a set of observations and recommendations surfaced at the November event representing a shared consensus (albeit not necessarily unanimity) across a broad array of invested communities in and around the UCSF community.

Observations and Recommendations

Topic 1: Governance & Accountability

As with the development and implementation of other complex processes, establishing structures for governance and accountability is consistently identified as a critical initial step in the ideal process of algorithm development and deployment.

Governance is the first step to getting to accountability, equity, and justice.
– Workshop Panelist

Recommendations

1. Establish an accountability structure (individual, institutional, etc.).
2. Establish sustainable mechanisms by which all recommendations in this report are assessed, implemented, and their status tracked, with consideration for agile governance in this deeply iterative space.

3. Ensure direct communication with invested communities, including dissemination of information in a variety of modes and formats to meet the needs of each.
4. Set standards and guidance for decision-making that address trade-offs, data use, enforcement, and redress of harms.
5. Build and actively maintain a registry of deployed algorithms that may impact equity, fairness, or justice; and establish clear criteria and a process for determining which algorithms require entry into the registry.
6. Provision certification and monitoring/audits.

See potential influencers and implementors in Appendix 1.

Topic 2: Community Engagement

An algorithmic justice approach entails reliable and consequential community engagement in all stages and facets of the development and deployment of institutional algorithms. Adequate representation and active inclusion of these collaborative partners—particularly representatives of communities that experience health disparities due to discrimination, inadequate access, and other structural barriers—are essential, as are the structuring of initiative plans and meeting agendas and language so that this collaboration might yield innovative and transformative approaches and practices. Done well, collaboration of this type has the potential to build meaningful trust between these communities and representatives of UCSF, ultimately resulting in UCSF creating more opportunities for collaboration and becoming a more responsible and broadly caring and just institution.

Recommendations

1. Analyze and provide opportunities for broad, equitable, and consequential community engagement in governance.
2. Provide opportunities for community engagement in model development and deployment, including advising on best practices for equity-framed model development and deployment, and routine inclusion of community voices in the implementation and evaluation of those models. Efforts should leverage expert scholarship in this arena as well as knowledge from current community engagement models at UCSF.

Similar to how UCSF has successfully incorporated community perspectives in other areas, we could envision a Community Advisory Board dedicated to algorithmic justice
– Workshop Steering Committee

3. Provide opportunities for community engagement in communication and dissemination efforts related to the development and deployment of models and ensure the appropriateness of language, terminology, channels, and platforms for communities bearing the burdens of health disparities.
4. Ensure demonstrable inclusion of broader community's core values/goals in any/all algorithm development processes, including the selection of performance metrics and continuous process improvement activities.

An important and meaningful way to sustainably engage a broad community is to nurture opportunities within the AI workforce
– Workshop Steering Committee

See potential influencers and implementors in Appendix 1.

Topic 3: Model Development & Deployment

Model development and deployment refers to the creation and implementation of algorithms, which includes data collection and preprocessing methods, selection of features and training methods, use of bias and fairness audits, assessments of impact and interpretability, and algorithm monitoring and maintenance. An algorithmic justice approach necessitates the application of ethics and equity considerations in each of these activities to ensure that models operate fairly and justly within the dynamic landscape of health care delivery options and patient preferences.

Recommendations

1. Create and disseminate guidance and technical standards for algorithm development and deployment that are aligned with reporting standards and includes specification of goals and intended target populations, expected impacts on health equity, training and testing data, and the basis for selecting metrics for optimizing performance.
2. Address data issues including diversity, quality, quantity, and data source(s); explicitly state potential limitations and sources of bias; and ensure that they are detected, monitored and mitigated prior to deployment.
3. Design models using approaches that prioritize health equity (e.g., weighting training data in proportion to the needs of different target subpopulations).
4. Ensure there is a “human in the loop” where appropriate and necessary, i.e., rather than relying entirely on automation, consider when and how to incorporate a person to oversee, evaluate, and validate the algorithmic development pipeline and deployment processes.

If health equity is an agreed upon goal, the resulting algorithmic tools will be different and will work toward decreasing health inequities
– Workshop Participant

See potential influencers and implementors in Appendix 1.

Topic 4: Model Reporting

Model reporting refers to communication of the nature of the samples (population characteristics, type of sample, etc.), the various measures (test results, demographic data, data annotation, caregiver notes, etc.), the mathematical methods and assumptions, and model performance metrics associated with algorithms and analytical models that infer hypotheses or specific conclusions in research or suggest specific clinical actions. Such reporting should reveal all analytical components underlying actionable outcomes to all interested parties.

Recommendations

1. Create Algorithm Fact Sheets that include how and why each algorithm was developed, performance metrics, quantification of uncertainty, potential sources of bias, and disclaimers related to intended use.
2. Develop a Dictionary of Terms (e.g., “fairness”, “equity”, etc.) and ensure cultural and language appropriateness in its design and dissemination.

See potential influencers and implementors in Appendix 1.

Topic 5: Education & Training

Because many aspects of data analytics are unfamiliar to experts in social justice, and *vice versa*, cross-training sufficient to establish transdisciplinary ‘working literacy’, if not expertise, is essential to achieving the full impact of algorithmic justice in precision medicine. Similarly, ‘lived experience’ of those impacted by algorithm-based clinical decisions is important to understand at multiple steps along the informational translation pathway. To that end, the workshop participants recommended broad education and training across the spectrum of invested communities.

The question about how we make decisions about tradeoffs is a fundamental ethical question. We need...education and engagement about the tradeoffs
– Workshop Panelist

Recommendations

1. Train model developers and users on health equity and health justice frameworks, the profound impacts of health and health care disparities on individuals and communities, decision-making trade-offs and their impacts.

2. Educate communities about AI development, use, and governance, as well as decision-making trade-offs and their impacts.

See potential influencers and implementors in Appendix 1.

Topic 6: Overarching Comments

In addition to the topically focused matters considered above, several general themes emerged that merit emphasis.

Trust in an algorithm isn't reliant on the algorithm itself but on the environment/institution within which it's used. Trust in UCSF is more important than trust in an individual algorithm in use at our institution.
– Workshop Participant

Overarching Observations and Recommendations

1. Just and equitable use of algorithms and analysis tools in health research and healthcare will require effective communication, a shared understanding of the value of these analysis tools, and the consideration of input from invested communities for any given research analysis or healthcare solution.
2. Workforce diversity at all levels must be fostered – i.e., in algorithm development algorithm deployment, and governance team.
3. Institutional commitment and funding are essential for implementation of each of the recommendations above.
4. It is essential to determine how algorithms can be used to uncover biases in the databases, algorithms, and related systems; and how algorithms can preclude such biases in establishing future databases.
5. Data management issues that broadly impact algorithmic justice such as data consent, control, and data “ownership” must be considered and addressed.

Appendix 1: Potential Influencers and Implementors

The committee that prepared this report has no authority to specify individuals or organizations at UCSF to be charged with addressing the recommendations offered in this report. However, to emphasize that accountability will require specific assumption of responsibilities, we present here one perspective on who could participate in implementing the proposals introduced here.

UCSF Organization	Report Topic				
	1	2	3	4	5
AI Ethics and Policy Task Force (IRB)	■				
Associate Chief Information Officer for Research (ACIO-R)		■			
Associate Vice Chancellor for Research (AVCR) - Inclusion, Diversity, Equity, and Anti-Racism (IDEA) Oversight Committee, Data and Technology Working Group	■	■	■		
AVCR-IDEA Oversight Committee, Community Engagement Working Group	■	■			
Center for Community Engagement Council		■			
Chancellor	■	■			
Chief Compliance Officer & Chief Privacy Officer	■				
Chief Data Officer	■				
Chief Health AI Officer for UCSF Health (CHAIO)	■				
Chief Information Officer (CIO)	■				
Chief Research Informatics Officer (CRIO)	■				
Committee on Enterprise Information and Analytics	■				
Clinical and Translational Sciences Institute (CTSI) Community Engagement Program	■	■			
CTSI Consultation Services	■				
CTSI Regulatory Knowledge and Support (RKS) Program	■				
Executive Vice Chancellor and Provost (EVCP)	■				
Impact Monitoring Platform for AI in Clinical Care (IMPACC)		■			
Institutional Review Board (IRB)	■	■			
Library			■	■	■
Office of Communications					■
Office Community & Government Relations	■				
Office of Diversity and Outreach			■		
Office of Ethics and Compliance			■		
Office of Research	■		■	■	
Patient and Family Advisory Council (PFAC) leadership - Patient Experience Office	■	■	■	■	
Restorative Justice Practices Program	■	■	■	■	
Schools of Medicine, Nursing, Dentistry, Pharmacy, and the Graduate Division	■	■	■	■	
UCSF Health Chief Executive Officer (CEO)		■	■	■	
UCSF Health Equity Council		■	■	■	■

Appendix 2: Glossary

Term	Definition	As used by...
Accountable	Accountable means being responsible or answerable for your actions, decisions, or duties.	General audience
	An individual or entity is legally responsible for their actions or conduct, and they can be held liable or answerable for any violations of laws, regulations, or legal obligations.	Legal community
Algorithm	An algorithm is a set of precise instructions or rules used to solve a specific problem or perform a particular task. Algorithms are fundamental in computer programming and are used to perform a wide range of tasks, from simple calculations to complex data analysis and decision-making processes.	General audience
Algorithmic Bias	Algorithmic bias refers to the presence of systematic and unfair discrimination in the outcomes produced by algorithms. Algorithms can exhibit bias when they result in inequitable or unjust treatment of different groups of people. This bias can arise from various sources, including biased data, biased design, or biased decision-making processes within the algorithm.	General audience
Algorithmic Justice	Algorithmic justice is a concept that focuses on ensuring fairness, equity, and accountability in the development and deployment of algorithms and artificial intelligence systems. It seeks to address and rectify issues related to bias, discrimination, and ethical concerns that can arise when algorithms are used to make decisions or automate processes in various aspects of society.	General audience
Artificial Intelligence (AI)	Artificial Intelligence (AI) refers to the simulation of human intelligence in computers and other machines. It involves the development of computer systems and software that can perform tasks typically requiring human intelligence. These tasks include problem-solving, learning, reasoning, understanding natural language, recognizing patterns, and making decisions.	General audience

Automated	Automated refers to a process, task, or system that is performed or controlled by machines, computers, or other technology without the need for direct human involvement.	General audience
Autonomous	Autonomous refers to the ability of a system, machine, or vehicle to operate and make decisions on its own, without direct human control or involvement.	General audience
Bias	Bias, in a general sense, refers to a tendency to favor or oppose something or someone unfairly or without full objectivity.	General audience
Bias	In an academic context, bias refers to a systematic and disproportionate influence or distortion in research, data collection, analysis, or interpretation that can lead to results or conclusions that are not representative of actual facts or underlying reality.	Scientific/technical community
Big Data	Big data refers to extremely large and complex sets of data that are beyond the capabilities of traditional data processing software and hardware to efficiently store, manage, and analyze. It encompasses vast amounts of structured and unstructured data, such as text, images, and sensor-generated information, often characterized by the three Vs: Volume, Velocity, and Variety.	General audience
Community	A community is a group of people who share common interests, characteristics, or goals and who interact with one another, cooperate, and form social connections within a particular geographic area or through virtual networks.	General audience

Community Engagement	Community engagement refers to processes of involving collaborators representing groups who should be served by the work but who often do not have access to it and/or are not part of its development or assessment.	General audience
Data	Data refers to information that can be collected, measured, or recorded. They can take many forms, including numbers, text, images, audio, or video. Data serve as the building blocks of information, which in turn creates knowledge.	General audience
Data Mining	Data mining is the process of assessing large amounts of data to discover hidden and valuable patterns, information, or insights.	General audience
Data-Driven	Data-driven is a term that describes a way of making decisions and taking actions based on information and evidence gathered from data. Instead of relying solely on intuition or opinions, a data-driven approach uses facts and statistics to drive conclusions.	General audience
Decision Making	Decision making is the process of choosing between different options or courses of action when faced with a problem or situation. The process typically includes gathering information, considering goals or objectives, weighing the pros and cons of each choice, and selecting an option that best fulfills needs or preferences.	General audience
Deep Learning	Deep learning is a type of AI that is inspired by the way our brains work. It involves training computer systems, often called neural networks, to learn and understand data in a way that is similar to how humans learn from experience.	General audience

Discrimination	Discrimination refers to unfair or differential treatment of individuals or groups based on certain characteristics or traits, such as their race, gender, age, religion, nationality, or disability, among others.	General audience
Disparity	Disparity refers to a significant difference or gap between two or more things, often in terms of characteristics, conditions, or outcomes. It indicates inequality or variation between groups or situations.	General audience
Electronic Health Records (EHR)	Electronic health records (EHRs) are digital versions of a person's medical history and health information. These records contain details about a patient's health, including their medical conditions, treatments, medications, test results, and doctor's notes.	General audience
Equity	Equity refers to fairness and justice in the way people are treated and the distribution of resources and opportunities. It ensures that each group or individual is provided an equal chance to succeed and reach their full potential, regardless of their background or circumstances.	General audience
Explainability	Explainability is the quality of being understandable and clear, especially with respect to complex or technical information, processes, or decisions.	General audience
FAIR Principles	The FAIR principles are a set of guidances that ensure that data are Findable, Accessible, Interoperable, and Reusable. They were developed to improve the management, sharing, and usability of data, particularly in the context of scientific research and data science.	General audience
Health Data	Health data refers to any information related to an individual's health, well-being, or medical history. This information can be collected and stored in various forms, including electronic health records, medical charts, laboratory results, patient questionnaires, wearable device data, and more.	General audience

Interpretability	Interpretability refers to perceiving the meaning of data or facts. Models are interpretable when humans can readily understand the reasoning behind predictions and decisions made by the model.	General audience
Justice	Justice is a complex and multifaceted concept that encompasses fairness, equity, and the principle of treating individuals or groups in a manner that is morally right and in accordance with the law. It involves upholding and applying laws and principles in a way that ensures that people receive what is due to them, whether it's protection of their rights, treatment, or entitlements.	General audience
Liable	Liable is used to describe a legal responsibility or obligation that an individual or entity has for their actions or conduct. Liability means legal accountability for a particular act or omission, conferring legal consequences, such as the payment of damages, penalties, or other remedies.	Legal community
Machine Learning	Machine learning is a subfield within AI that focuses on developing algorithms and models that enable computers to learn and make predictions or decisions based on data. In other words, machine learning is a key technique used to achieve AI's broader goal of intelligent behavior.	General audience
Natural Language Processing (NLP)	Natural language processing (NLP) is a branch of AI that focuses on enabling computers to understand, interpret, and interact with human language in a way that's similar to how people communicate. In other words, NLP helps computers understand and work with human language.	General audience
Precision Medicine	Precision medicine employs advanced computing algorithms to aggregate, integrate and analyze data from basic and translational science, clinical, personal, environmental and population health settings, to better understand biological processes and define disease mechanisms, and to predict more precise diagnostics, therapeutics, and prevention measures.	General audience

Prediction/Predictive Algorithm	Prediction is the process of forecasting or estimating future events or outcomes based on available information, data, patterns, and analysis of past and current trends. A predictive algorithm, also known as a predictive model, is a technique used in data analysis and machine learning to make predictions.	General audience
Structural Racism	Structural racism refers to a system of societal, economic, and political structures, policies, and practices that systematically and disproportionately disadvantage certain racial or ethnic groups while benefiting others. It is a form of racism that is deeply ingrained in the fundamental institutions and systems of a society, perpetuating inequities and discrimination over time.	General audience
Underrepresented	Underrepresented refers to a situation where a particular group or category of people is not adequately or proportionately represented in a specific context, organization, or field relative to their presence in the general population or their potential contribution.	General audience
Underserved	Underserved refers to a condition in which a particular group or community does not have adequate access to essential services, resources, or opportunities that are considered necessary for a decent quality of life or equal participation in society. These services and resources may include healthcare, education, housing, economic opportunities, or other vital social goods.	General audience