

LITIGATING ALGORITHMS:

CHALLENGING GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS

An AI Now Institute Report

In collaboration with

**Center on Race, Inequality, and the Law
Electronic Frontier Foundation**

SEPTEMBER 2018

AINOW

TABLE OF CONTENTS

INTRODUCTION

WORKSHOP SUMMARY

FIVE EXEMPLARY ARENAS OF LITIGATION

KEY QUESTIONS

SESSION ONE: MEDICAID AND GOVERNMENT BENEFITS

SESSION TWO: PUBLIC TEACHER EMPLOYMENT EVALUATIONS

SESSION THREE: THE ROLE OF SOCIAL SCIENCE AND TECHNICAL EXPERTS

SESSION FOUR: CRIMINAL RISK ASSESSMENT

SESSION FIVE: CRIMINAL DNA ANALYSIS

AREAS FOR FUTURE DISCUSSION AND COLLABORATION

INTRODUCTION

There is currently much debate over the use of algorithmic decision systems in our core social institutions. From criminal justice to health care to education and employment, we are seeing computational and predictive technologies deployed into or supplanting private and governmental decision-making procedures and processes. As a result, many advocates, academics, and policymakers have begun to raise concerns, urging adequate safeguards, oversight, appeal, and redress mechanisms for protecting vulnerable populations from harm. For example, New York City recently established the first city-wide [Automated Decision System Task Force](#) to study and recommend policies, practices, standards or other guidelines on the use of such systems across all of its public agencies.

In service of better understanding how these systems affect us and what strategies are being used to challenge their unforeseen or unlawful outcomes, AI Now recently teamed up with [NYU Law's Center on Race, Inequality, and the Law](#), and the [Electronic Frontier Foundation](#) with the support of the [MacArthur Foundation](#) to host a workshop at NYU Law. This workshop focused on conducting a deep examination of current United States courtroom litigation where algorithmic decision making has been central to the rights and liberties at issue in the case. Throughout the day participants reviewed cases including: [K.W. v. Armstrong](#), where the ACLU of Idaho challenged the state's use of a black box algorithm to make determinations for its Medicaid program; [Houston Federation of Teachers v. Houston Independent School District](#), where a public school teachers union challenged the use of proprietary algorithms for school employment practices; and a case from a Washington, D.C. juvenile criminal court that rejected the use of a Violence Risk Assessment system because it lacked sufficient scientific merit under the Supreme Court's [Daubert](#) standard for expert testimony in court. This workshop was the first in the United States to critically examine litigation strategies for challenging government use of algorithmic decision systems across different disciplines and areas of law. Below is our summary of the event, including links to key court documents and recommended areas for future discussion and collaboration.

WORKSHOP SUMMARY

In June 2018, we convened the workshop with the goal of bringing together legal, scientific, and technical advocates who are focused on litigating algorithmic decision-making across various areas of the law (e.g. employment, public benefits, criminal justice). We structured the day with the practical aim of discussing strategy and best practices, while also exchanging ideas and experiences in litigation and other advocacy in this space. The gathering included several of the lawyers who brought the cases alongside advocates, researchers, technical experts, social scientists, and other leading thinkers in the area of algorithmic accountability.

We began the day with a broad overview and framing of the issues, noting that while many papers, reports, and conferences have examined legal and policy issues involving algorithmic decision systems, most focused solely within distinct areas of law, such as labor, criminal justice, education or health, with few comprehensively examining the entire legal landscape. Nor had there been the opportunity to focus exclusively on the tactics, challenges, and practice of litigating cases of this type. The discussion addressed methods of discovering evidence to show how and when legal rights have been violated; approaches to framing the facts and theories of a legal complaint; what strategies have the greatest influence on the directions and decisions in a given case; what remedies to seek; and other common themes and lessons that have emerged throughout the legal process.

Workshop attendees came from various advocacy, policy, and research communities, including NYU Law, ACLU National, NYC's Legal Aid Society, Legal Aid of Arkansas, Innocence Project, Civil Rights Corps, National Health Law Program, ACLU of Idaho, Federal Defenders of New York, New York Civil Liberties Union, Center for Constitutional Rights, Public Defender Service for the District of Columbia, Yale University, NAACP Legal Defense and Educational Fund, Columbia University, the National Association of Criminal Defense Lawyers, Microsoft Research, and Upturn.

FIVE EXEMPLARY ARENAS OF LITIGATION

To provide a representative sense of the scope and scale of current litigation involving algorithmic decision systems and their impact on rights and liberties, we structured the day into five sections, each focused on specific areas where we anticipated seeing more activity and conversation over time. These were:

1. Medicaid and Government Benefits
2. Public Teacher Employment Evaluations
3. The Role of Social Science and Technical Experts
4. Criminal Risk Assessment
5. Criminal DNA Analysis

For the sessions focused on specific cases, we asked the lawyers and experts to walk us through their cases and issues in terms of process, strategy, key decisions, and lessons learned, before opening discussion up to the room. We also asked each presenter to share specific legal or technical documents that could help participants gain insight into the cases, including legal complaints, expert reports, court orders, and other materials. The third session brought together leading scientific and technical experts to discuss their roles in shaping legal proceedings and informing the outcomes.

KEY QUESTIONS

To frame the conversation, we posed a series of prompts for each presenter:

- How did your case/issue emerge? Who brought it to your attention?
- Why did you decide to engage in litigation versus prioritizing other strategies, such as community activism, policy advocacy, media attention, etc.?
- What was the legal theory behind your case?
- How did you choose whom to sue and where?
- What types of factual and legal investigations did you undertake to develop the basis for your case?
- Did you engage specific technical, social science, or legal experts before you sued? Or after? Who were they? What was their expertise?
- How did you gather information on the algorithm in question? Open Records Requests? Individual interviews? Court-sanctioned discovery?
- What remedial options were available to your client? Did you get what you sought, or did you have to engage in additional litigation to do so?
- What would you have done differently? What would you have changed about your litigation strategy?

SESSION ONE: MEDICAID AND GOVERNMENT BENEFITS

- **PRIMARY PRESENTERS:** Ritchie Eppink, *Legal Director, ACLU Idaho*; Kevin De Liban, *Economic Justice Workgroup Leader, Staff Attorney, Legal Aid of Arkansas*; Elizabeth Edwards, *Senior Attorney, National Health Law Program*
- **MODERATOR:** Jennifer Lynch, *Senior Staff Attorney, EFF*
- **KEY DOCUMENTS**

This session explored the intersection of health, public benefits, administrative, and Constitutional law through the lens of several cases challenging the use of algorithmic decision systems to determine how, when, and what amount of government benefits individuals would receive. Specifically, we looked at cases involving health and disability benefits. In these cases, faulty algorithmic decision systems improperly diminished or terminated benefits and services to individuals with intellectual, developmental, and physical disabilities. What emerged was a clear vision of new challenges and risks for these vulnerable populations.

The problems appear to emerge when agencies that oversee these benefits adopt complex yet archaic algorithmic formulas to govern benefit disbursement and patient assessment. These technical systems are then often implemented without any meaningful training, support, or oversight. Nor is their use accompanied by protections for or involvement of the benefits recipients they are intended to serve.

Generally, government agencies that adopt algorithmic decision systems justify them as measures to produce cost savings and standardization. Accordingly, these systems typically target populations that are considered the “most expensive,” which often include the most politically, socially, and economically marginalized communities, who because of their status, are more likely to need greater levels of support. Yet, the agencies often fail to conduct a rigorous assessment of the true costs and benefits of such systems, particularly on the people who will be most affected by their implementation. Many states simply pick an assessment tool used by another state, trained on that other state’s historical data, and then apply it to the new population, thus perpetuating historical patterns of inadequate funding and support. In addition, there are frequent flaws and errors in how these assessment systems are implemented and in how they calculate the need for care. Government agencies adopting these systems commonly enter into contracts with third-party vendors that handle everything. The agency, particularly front-line staff who are most familiar with the Medicaid program and its challenges, have little to no involvement

in how the tool analyzes data and produces calculations. Because these tools are private systems licensed to government agencies, the design specifications and particularities of the technical system are often considered trade secrets of the vendor and are not publicly available.

We also heard several examples where algorithmically-driven public benefit assessments relied too heavily on quantitative data, such as financial costs and number of hours of care, and failed to take into account more qualitative information, such as the benefits of remaining at home and of being integrated into one's community, which serve both the patient and save the state money in the long term. Algorithmic systems also tend to overestimate the extent to which a person's family or other community support are capable of providing services in the absence of public benefits.

Algorithmically-driven benefit determinations also affect labor practices, which in turn affects the quality of care. For instance, disability benefit allocations have traditionally been driven by human nursing staff conducting individual one-on-one assessments of each beneficiary. When this is shifted to automated "black box" data calculation processes, it eliminates any meaningful dialogue between the beneficiary, a provider, and the state about their care and condition needs. Such dialogue may often result in better data and creative solutions for care. Additionally, algorithmic decision systems tend to overemphasize certain conditions, while not fully accounting for others, despite similar care needs. For instance, most algorithmic decision systems used in healthcare fail to identify or distinguish cognitive and mental health issues, which can result inadequate or inappropriate care and benefits allocations.

In terms of litigation outcomes, we heard about several early victories in these cases based, in part, on constitutional and administrative due process claims that challenged the lack of notice, explanation, and ability to comment or contest the changes to public benefit systems. This was especially relevant for the plaintiffs, individuals with intellectual or developmental disabilities. However, as systems become more widely adopted and accepted across different jurisdictions and domains, and affect different demographics, these challenges may be harder to bring, at least on grounds that challenge lack of notice. It is also worth noting that these challenges and the resulting mitigation efforts are extremely resource intensive, which can serve as an additional barrier for advocates.

Procedural due process arguments were particularly adept at dismantling claims that these algorithmic decision systems are proprietary trade secrets, especially when used for public benefits determinations. The public's right to know how their benefits are being determined is a strong case against trade secrecy. Procedural due process also appears well-suited for challenging algorithms that fail to adjust or account for administrative rule changes. For example, if a new rule went into effect altering the obligations of family members to provide services for

adults with certain disabilities, algorithmic systems that continued to rely on the old rule would likely be open to challenge.

Another successful basis for challenging algorithmic decision systems was to highlight errors in the software design or implementation. Note that this is predicated on having access to technical information about the system and access to experts who have the ability to review and interpret the system, both of which can be difficult to obtain. However, if given access with adequate resources to interrogate the systems, such challenges can be productive. For example, in Arkansas, an algorithmic decision system allocating home health care to Medicaid patients failed to accurately understand the care needs of patients with cerebral palsy or diabetes. Yet, this was only discovered during the course of litigation, and only after the system's code and its associated technical documentation had been carefully examined. Transparency and public access should remain important goals for advocates interested in litigating algorithmic decision systems because similar errors likely exist and could be integral to future challenges.

Finally, we learned that once a system is found to be discriminatory or otherwise inaccurate, there is an additional challenge in redesigning the system. Ideally, government agencies should develop an inclusive redesign process that allows communities affected by algorithmic decision systems to meaningfully participate. But this approach is frequently met with resistance. Currently, the plaintiffs in the Idaho case are working with the Idaho Department of Health and Welfare to develop a new model. We hope the outcomes of that process can be instructive for future system redesigns.

SESSION TWO: PUBLIC TEACHER EMPLOYMENT EVALUATIONS

- **PRIMARY PRESENTER:** Martha Owen, *Partner*, Deats, Durst, Owen & Levy, P.L.L.C.
- **MODERATOR:** Jason Schultz, NYU Law | AI Now
- Key Documents

The second session explored one of the few successful cases challenging the use of proprietary algorithms to evaluate the performance of public employees. To guide us through the case, we heard from Martha Owen, lead counsel for the Houston Federation of Teachers (HFT) in their suit against the Houston Independent School District. Her story was similar to those involving public benefits determinations, in that they both faced barriers in the attempt to discover how and where such systems affected individuals. In Houston, the system began as a method for determining bonuses for public school teachers by evaluating student performance on standardized tests. Later, the school district began applying a similar algorithmic methods to sanction employees for low student performance on standardized test. When HFT members asked to examine the systems, they were denied with the explanation that the algorithms and code that comprised these systems were the private property of a third-party vendor. HFT then brought suit on several different grounds, ultimately winning the right to go to trial on a constitutional procedural due process claim that focused on the school district's failure to give the teachers meaningful access to the algorithmic system involved in the district's employment decisions.

The discussion provided many insights into algorithmic evaluations of public school teachers and the ways in which teachers and their unions are fighting to ensure such methods are fair, accountable, and sufficiently transparent. In this case, the third-party software company fought to keep its source code, training data, and design as secret as possible, which gave litigators an opening to use procedural claims.

The role of experts was critical, both within the litigation and as part of building a broader consensus that the algorithmic system was inherently flawed and unreliable as a predictor of teacher performance. The vendor only allowed HFT's expert to review the system under extreme constraints. Under this limited review, the expert found that the algorithmic system was uninterpretable, which bolstered the procedural due process claims. By combining evidence of procedural due process problems and faulty scientific methodologies, the teachers were able to mount a strong and ultimately successful case.

SESSION THREE: THE ROLE OF SOCIAL SCIENCE AND TECHNICAL EXPERTS

- **PRIMARY PRESENTER:** Harlan Yu, *Executive Director*, Upturn; Ashkan Soltani, *Technology Expert*; Aaron Pallas, *Chair*, Department of Education Policy and Social Analysis, Columbia Teachers College; Hanna Wallach, *Senior Researcher*, Microsoft Research NYC; Daniel Kahn Gillmor, *Senior Staff Technologist*, ACLU Speech, Privacy, and Technology Project.
- **MODERATOR:** Rashida Richardson, *Director of Policy Research*, AI Now

In this session, we asked a panel of social science and technical experts to discuss the work they do to help lawyers, judges, and jurors understand the methodologies and technologies at issue in these cases. They specifically highlighted best practices for working in multidisciplinary teams and ideas for building out expertise across different disciplines and areas of practice.

One important finding from the conversation was the necessity of analyzing the “model” or framework of assumptions made when developing and implementing an algorithmic decision system that attempts to represent the real-world context in which the system will operate, instead of simply examining the algorithm that operationalizes these assumptions as part of one or more systems. For example, in the Arkansas public benefits case, the framework itself failed to diagnose or identify cerebral palsy and diabetes. The implementation of the framework as software also included several coding errors. Each is a distinct point of interest when litigating algorithms.

Other experts noted that when we talk about racial bias or other social problems reflected in algorithmic systems, we are often talking about problems with the framework itself; the attempt to represent the system’s operational context. This can either fail to reflect the conditions in which the system will be used or oversimplify systemic and structural biases that are already present in existing real-world processes. For example, relying solely on student test results as a measure of teacher performance both fails to acknowledge alternate metrics of learning and success, and fails to account for school, classroom, and socioeconomic factors that significantly affect performance, such as classroom size and language barriers. These errors and oversights are particularly prevalent in frameworks developed without direct involvement of those who will

be affected by a given system, who can help illuminate the full range of factors that may influence the context the framework intends to represent.

Experts also expressed concern that the emphasis on faulty algorithmic decision systems shifts attention away from big picture, broader advocacy. Most public agencies struggle with a lack of adequate resources (e.g. underfunding or understaffing), and the unfortunate reality of this resource deficiency is that some communities are and will continue to be underserved or suffer other negative consequences. While algorithmic decisions systems can distort and exacerbate these problems, they also tend to mask or legitimize the policy decisions that create them. In particular, the experts noted that some of the outcomes in the public benefits cases could have been mitigated sooner if the system's outputs were initially tested. They posit that policymakers have a hard time determining what the optimal outputs should be, so they look to technical solutions that allow them to avoid these difficult policy decisions.

SESSION FOUR: CRIMINAL RISK ASSESSMENT

- **PRIMARY PRESENTER: Blase Kearney**, *Supervising Staff Attorney, Trial Division, Public Defender Service for the District of Columbia*
- **MODERATOR: Vincent Southerland**, *Executive Director, Center on Race, Inequality, and the Law, NYU*
- **Key Documents**

This session provided a detailed account of a juvenile criminal sentencing hearing in Washington D.C., where the child defendant's lawyer was able to secure a ruling from the judge that a long-standing "Violence Risk Assessment" system had not been sufficiently validated to be admissible as part of expert testimony in the case.

The presenter explained that these violence risk assessment systems have a powerful influence over criminal sentencing outcomes, especially for children. While a young person in the criminal justice system may demonstrate, through their pre-sentencing behavior, that serious criminal sanctions are unnecessary, a "high risk" finding on one of these algorithmic assessments can result in them being sent to a psychiatric hospital or a secured detention facility, separating them from their family and drastically changing the course of their life. Moreover, young people often plead guilty to violent offenses on the condition that they will be eligible for probation rather than incarceration, if they comply with certain court requirements including algorithmic risk assessments. When the risk assessment produces a high risk score, that score changes the sentencing outcome and can remove probation from the menu of sentencing options the judge is willing to consider. Thus, the civil liberties of a young person can often depend entirely on the outcome of the risk assessment system.

In examining these systems, many advocates have raised significant concerns about embedded racial bias. For example, most assessment systems include several risk factors that function as proxies for race. One risk factor that is often used is "parental criminality" which, given the long and well-documented history of racial bias in law enforcement, including the over-policing of communities of color, can easily skew "high risk" ratings on the basis of a proxy for race. "Community disorganization" is another influential risk factor if an individual lives in a neighborhood considered to be "violent" or near gang activity, which given the long and well-documented history of private and public housing discrimination, could skew "high risk" ratings on the basis of a proxy for race.

We also learned that these systems are rarely challenged. For example, the system in the DC case had been in use in the juvenile criminal system since 2004, but had not been challenged until 2018. Criminal defense participants suggested that most defendants and their lawyers at the trial level simply do not have the time, energy, or expertise to raise such challenges. And even when more resourced groups like the Public Defender Service of DC bring challenges, the system is designed at every step to keep litigants from getting to the information needed to raise the issues appropriately.

Fortunately, in this particular case, the defense attorneys were able to get access to the questions used to administer the risk assessment as well as the methods of administering it. When they dug into the validity behind the system, they found only two studies of its efficacy, neither of which made the case for the system's validity; one was 20 years old and the other was an unreviewed, unpublished Master's thesis. The long-held assumption that the system had been rigorously validated turned out to be untrue, even though many lives were shaped due to its unproven determination of 'risk'. With this in hand, attorneys were able to raise a *Daubert* challenge and succeed, obtaining a ruling that the risk assessment was not sufficiently validated to be admissible.

SESSION FIVE: CRIMINAL DNA ANALYSIS

- **PRIMARY PRESENTERS:** **Jessica Goldthwaite**, *Staff Attorney*, Legal Aid Society (NYC) - DNA Unit; **Chris Flood**, *Federal Public Defender Office*, NYU; **Nathan Adams**, *Systems Engineer*, Forensic Bioinformatics
- **MODERATOR:** **Vincent Southerland**, *Executive Director*, Center on Race, Inequality, and the Law, NYU
- **Key Documents**

In this session, we heard from several criminal defense attorneys and experts who have litigated cases involving the prosecution's use of probabilistic genotyping software to identify alleged perpetrators based on DNA evidence gathered from the scene.

A critical point raised by the presenters was that the algorithms used by such systems are now so complex that most medical examiners and laboratory technicians are unable to replicate the computational results without the assistance of the system. In other words, they cannot replicate the results simply through human effort and analysis alone.

We also learned that while DNA laboratories are often tested and certified to ensure they maintain minimum standards for biological or chemical testing, the systems they use to perform the probabilistic genotyping are often untested, especially for bugs in their code. This discrepancy between standards for biological science and computer science raised important concerns and points of further inquiry for future litigation.

This disciplinary distinction also emerged in the discussion of ways defense attorneys can challenge DNA analysis software under the *Daubert* standard for courtroom admissibility. In particular, the *Daubert* standard requires prosecutors to show that the relevant scientific communities that research and use the testing technology have generally accepted it as a valid method of testing for the application at hand. For DNA analysis systems, this would include molecular biologists as well as computer scientists and software engineers. It might also include scientific communities that study and seek to improve algorithmic accountability. This discussion generated interest in bringing these scientific communities into the debate over admissibility for all algorithmic systems.

AREAS FOR FUTURE DISCUSSION AND COLLABORATION

Many important lessons were learned in the course of the workshop. Below are several of the most prominent, including ideas and aspirations for future work.

- Exploring what an “advocacy-based algorithm” might look like that is oriented around the needs of individuals or communities that are impacted by these systems. The re-design process in the Idaho case may be instructive in exploring this approach.
- Developing a broader advocacy strategy that integrates concerns over algorithms with broader social, economic, and political concerns, such as the fact that agencies are underfunded and understaffed, teachers are underpaid, and low-income communities and communities of color are overpoliced. How do we make sure we are addressing the systemic and structural concerns in these areas, and not just tinkering with the algorithmic systems at the margins?
- Asking more questions about the software written to implement algorithmic systems, the underlying algorithms, and the data they use. Software implementation bugs can lead to significant errors. There is often an assumption that errors in software design can be easily caught or fixed, but this is not true for several of the systems discussed at the workshop.
- Ensuring there is a greater emphasis on robust assessment of these systems before and during implementation, and bringing in external auditing when needed. This includes making sure standards of assessment are followed even when algorithms are procured from third-party vendors. It also means ensuring that information is not lost or manipulated, and there are consequences in place in the event that information necessary for litigation is not retained.
- Better recognition of the multidisciplinary nature of the problems raised by these systems, and acknowledgement of the way these systems move between legal rights and areas of concern that cannot be litigated.

- Highlighting the conflict between procedural due process rights and claims of proprietary or trade secrecy status for these systems upfront, so that decision makers understand the inherent tension and incompatibility between them.
- Improving the conversation between advocates and the communities that are affected by these systems. How do we work with these stakeholders in ways that center their voice and experience as expert, and ensure they are empowered to speak to the issues these systems raise? Also how do we figure out what affected individuals or communities need to know to meaningfully engage?