Data-driven discrimination: a new challenge for civil society

*Data-driven technologies have been a transformative force in society. However, while such innovations are often viewed as a positive development, discriminatory biases embedded in these technologies can serve to compound problems for society's more vulnerable groups. Having recently published a report on automated discrimination in data-driven systems, **Jędrzej Niklas** and **Seeta Peña Gangadharan** explain how algorithms discriminate, why this raises concerns for civil society organisations across Europe, and what resources and support are needed by digital rights advocates and anti-discrimination groups in order to combat this problem.*

In recent years, debate on algorithms, artificial intelligence, and automated decision-making has stoked public concern, panic, and occasional outrage. While such innovations are very often shown in a positive light, there are also stories of vulnerable groups who struggle because of discriminatory biases embedded in the technologies. More often than not, public discourse presents these problems in a distinctive US context. In our new report "Between Antidiscrimination and Data: Understanding Human Rights Discourse on Automated Discrimination in Europe", we make visible European perspectives on data-driven systems through the lenses of 28 civil society organisations (CSOs) active in the field of human rights and social justice in nine EU countries.

How do algorithms discriminate?

We began our study by reviewing the problem of algorithmic or data-driven discrimination. In a very broad sense, algorithms are encoded procedures or instructions. They often use data as their main ingredient (or input), transforming these inputs into a desired output, based on specific calculations. Automated systems based on algorithms are complicated and vary in character, purpose, and sophistication. The variety of systems also means that algorithmic discrimination can arise for various reasons.

To run, algorithms need data. But data can be poorly selected, incorrect, incomplete or outdated, and can even incorporate historical biases. One of the early examples (1988) of this problem was the case of St. George's Medical School in the UK. An automated system was used to screen the incoming applications from potential students. Modelled on previous job recruitment data, the system incorporated historical biases in its analytical process and discriminated against women and people with non-European names.

Concerns not only relate to the quality of input data but also extend to the design of the algorithm that is using those inputs. Programming decisions are essentially human judgements, and reflect a vision about how the world ought to be. For example, humans must decide on error types and rates for algorithmic models. In other words, someone has to decide whether to measure the algorithmic "reliability" in terms of the cases wrongly included in an algorithmic decision (e.g. false positives) or wrongly excluded (e.g. false negatives) from an analytic model. Someone also needs to decide what an acceptable level of wrongful inclusion or exclusion might be.

Differences over the setting of thresholds for determining an accurate misclassification will have significant consequences in determining what counts as discrimination in data-driven or algorithmic decisions. As discussed at a recent workshop entitled "Intersectionality and Algorithmic Discrimination" and explored at length in a paper by Michael Kearns, one of the best examples of such a problem was the widely discussed case of a sentencing algorithm used in the USA. The system assesses the risk of a defendant committing more crimes in the future, which affects the decision that determines a defendant's temporary detention. ProPublica journalists found that that system may treat black people less fairly than their white counterparts: "The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants", while "white defendants were mislabeled as low risk more often than black defendants".

Oppression produced by data

But discrimination through data doesn't need advanced technologies. In our interviews, we discovered that, by and large, CSOs worry less about automated discrimination and more about data and discrimination. There is a long tradition of creating separate registers and databases of marginalised communities, and our interviewees are concerned, working to stop their creation, or fighting to mitigate their impacts. In Sweden, for example, the police had been using an electronic registry containing data on people of Roma origin. Action taken by one of our participant CSOs in this study – Civil Rights Defenders – led to two court rulings admitting that this data practice was discriminatory, and that affected individuals must be compensated. We heard of similar problems in Poland, where anti-poverty organisations successfully opposed the idea of creating a central database of homeless people.

The link between data and discrimination is not limited to oppressive databases. In fact, many of our participants advocated the collection of information about minorities, in order to measure the state of social inequality. They see information about discriminated groups in many cases as a "missing link" that can be mobilised to guarantee real equality. Very often specific communities and their struggles are invisible in official datasets. For example, the Romanian census undercounts the actual number of Roma people living in the country by an estimated one million people. The distorted picture of the Roma population informed authorities' public policymaking, including the allocation of appropriate budgetary resources. Good and detailed knowledge of specific groups in society may be an essential tool to plan, implement, and evaluate public policies that are fair to all.

Some CSOs also stressed that some data-driven systems don't take into account the needs of specific communities. The design of national ID number systems has created struggles for both transgender people and migrants to gain access private and public services. For example, in Sweden, when a person undergoes gender reassignment they receive a new personal identity number which contains a new gender marker. Any public and private services tied to the ID system, such as banking, healthcare, and insurance, treat the newly ID'd individual as a completely different person. Without any connection to past transactions, interactions or records, the transgender person experiences disruption or denial of past services, many of which are critical to their daily survival.

A job for nerds

As data-driven decision-making becomes more central to everyday life, our interviewees are daunted by the complexity of algorithmic systems. This should come as no surprise. Very often, unequal treatment that is a consequence of data processing or algorithms will be difficult to detect. Algorithmic systems function as "black boxes": people might not even know that a machine has made a decision about them. Even if they did know, they (as well as data scientists) have limited abilities to understand the logic of how the decision was made. This is due to both technological constraints and also legal limitations (e.g. trade or professional secrecy) that prevent information being made available in the public domain.

Yet, whether knowing or unknowing about the problem of black boxes, human rights organisations feel disquieted and ill-equipped to intervene on behalf of affected populations. One participant from Poland commented that new problems created by digital technologies compel the creation of "job[s] for nerds": competent staff are needed to discover problems and opportunities related to the advent of automated systems.

But current uptake or involvement in the issue of automated discrimination is low. In our study, very few organisations had experience of working on the issue of algorithmic systems. A handful of CSOs highlighted the potential harms for specific vulnerable populations, while others focused on the broader problem of transparency in automated systems. Some interviewees pointed to problems like the possibility of errors creeping in, and the dehumanisation of the decision-making process, particularly in the context of automated welfare systems. But while evidence of some familiarity could be found in this handful of interviewees, direct advocacy — whether legal or technical or some combination of both — remains elusive for those with whom we spoke.

Next steps

An easy way to read CSOs' lack of involvement in data-driven discrimination is in terms of tech savviness—or the lack thereof. Indeed, the discriminatory impact of data-driven technologies is complicated and requires the civil society sector to think and act on more sophisticated technical knowledge. But we see great potential in CSOs evolving new and non-standard approaches to the problem of automated discrimination based on their diverse and sometimes competing approaches to dealing with data collection and marginalised populations.

CSOs can begin to confront automated discrimination by applying non-discrimination principles to data protection frameworks. In Europe, the General Data Protection Regulation (GDPR) creates some promising potentialities. Here the data protection impact assessment, special regime of sensitive data and safeguards in algorithmic decision-making may be called on. However, there are also other tools that can be used to bring more transparency and accountability to automated systems. For example, the AI Now Institute based at New York University has developed a toolkit which can help affected communities and stakeholders assess the use of AI and algorithmic decision-making in public agencies and determine where – or if – their use is acceptable.

We also heard our interviewees articulate a path forward, including solutions that acknowledge that the complex problem of automated discrimination requires a complex ecosystem. Some even mentioned the need for journalists to bring their investigative skills to expose data-driven systems. Their engagement with whistleblowers, as well as their skill in conducting corporate and public research, may serve as a base for further advocacy by CSOs. There are some good examples here, most notably various investigations by ProPublica. In view of the shrinking world of investigative journalism, human rights groups can also serve as companion investigators.

Generally we see three ways to support CSOs in engaging with the problem of automated discrimination:

- Resource digital rights or data privacy advocates to recognise anti-discrimination as a key concern for data protection, and undertake automated discrimination as a priority for their work
- Support anti-discrimination groups and other groups focused on equity and justice in recognising connections between their core work and values and "hi-tech" discrimination
- Acknowledge, cultivate, and support a flexible approach to highlighting and problem solving for automated discrimination.

This blog post originally appeared on the LSE Media Policy Project blog and is republished with permission.

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

About the authors

Jędrzej Niklas *is a Research Officer for the Justice, Equity and Technology (JET) project at LSE.*

Seeta Peña Gangadharan *is Assistant Professor in the Department of Media and Communications at LSE.*