EVENT NOTE

Workshop
Automation, Prediction, and Digital Inequalities
26 April 2016

The workshop was held under the Chatham House rules. This half-day event brought together academics from the fields of media and communications, computer science, journalism and public policy, as well as data practitioners, civil servants and civil society advocates. This event note, prepared by the LSE Media Policy Project, serves as a record of the core points of the discussion. It is not a verbatim summary nor is it a statement of a consensus position.

General overview

The main objective of the workshop was to explore the question of fairness in big data systems. Big data systems - also referred to as automated computer systems that analyse data in order to make predictions or determine outcomes - can produce biased or discriminatory outcomes that disfavour a particularly group, including people with protected characteristics, such as racial and ethnic minorities or women. This bias may result from a variety of sources, including non-representative data, data that essentially serves as a proxy for protected characteristics, or flawed analytic or algorithmic techniques. Such bias can result in certain groups adopting particular behaviours, such as buying insurance with unfavourable terms, or being targeted in particular ways, such as being persistently watched by police authorities.

The three key practical goals of the workshop were to:

- investigate new research trends around the notion of “digital inequalities”;
- discuss challenges of automation and prediction for various areas of everyday life; and
- identify potential harms of data-driven discrimination.

The workshop was divided into two sessions: the first on the value of data; and the second on the means of confronting discrimination in automated systems. Each session was structured around provocations from 3-4 academics or practitioners on a particular theme linked to their research, to which the group responded. The idea was to provoke debate around issues such as fairness, justice, inequality, uncertainty and responsibility, in the context of predictive analytics and big data.

The discussion revealed that people have a range of views about what counts as personal data given the precision of analytic techniques in identifying users from seemingly impersonal digital information, and, more importantly, about the kinds of value that is attributed to data. Many participants expressed concerns about the unique characteristic of predictive analytics, which includes the re-use of data/datasets for a purpose for which it was not originally intended. Some examples of this, each of which were discussed in detail at the workshop, include: the use of social media data by the police for monitoring/behavioural prediction purposes; the use of consumer data in the

allocation of labour resources, including workplace monitoring of employees; and the use of socio-demographic data (such as age, gender and so on) by employers when making hiring decisions about individuals.

The Overarching Questions

Provocateurs in the first session focused on exploring the value of data from various economic, sociological, cultural and historical perspectives. The following topics were the focus of the four provocations: the distribution of welfare in data privacy "markets"; police use of predictive analytics and social media surveillance as a means of responding to public protests; the relationship between workplace analytics and employee surveillance; and historical changes in retail pricing and consumer targeting.

Provocateurs in the second session considered technical solutions, market-based solutions, and existing legal frameworks as means by which incidents of data-driven discrimination could be addressed. Discussion ranged from a need to address how (and how well) technologists design and develop machine-learning algorithms, to the extent to which the public ought to trust the market to identify technology products and services with potentially questionable analytics practices. There was also a discussion of the lack of consensus about a way forward (not only within the data protection community, but also between data protection and human rights communities) that prevent meaningful engagement with the problems of data-driven discrimination.

Collectively, the provocations raised common questions related to automated computer systems: who benefits from such systems being in place? How are the costs of such systems distributed? What are the possibilities for resisting how personal data is being used? How should responsibility be borne, and by whom? What normative ideals of fairness and equality shape how we understand data discrimination problems and solutions?

The Issue of Black Box and Public Engagement

Two main issues in relation to predictive analytics surfaced during the workshop.

1. First, participants enquired about the general public's knowledge of, and attitudes towards, the issues of big data and predictive analytics. At present, the public's attitude can perhaps be summarised as disengaged, which some participants noted should not be surprising given the complexities of the phenomena in question. While some participants maintained that market self-regulation would be sufficient to address possible abuses of power related to predictive analytics (see section below), others believed that better informing the general public about how such systems work, and what users' rights are, would be more appropriate. There was no consensus about which approach might be most effective.

2. Second, the "black box phenomenon" (i.e. our inability to reverse engineer predictive algorithms and thus to understand how a computer programme generated a particular outcome or decision, or replicated the same results over time) was given special attention. Participants raised concerns over a lack of access to predictive systems (i.e. which would enable them to analyse and assess algorithms). There were also concerns about the commercial and closed nature of such systems: individuals and organisations that depend on proprietary automated systems or

software often have neither full control nor complete understanding of their inner workings. The participants did not reach a consensus on how better to govern, regulate or implement such systems, but had particular concerns about a general lack of accountability.

Regulation

According to some participants, a new regulatory regime, including in Europe, would require close collaboration between social sciences, legal studies, and computer science to understand both the processes and impact of new data analytics tools. Some participants highlighted a need to avoid the dominance of a US perspective when examining data-driven discrimination. One possible alternative is to encourage non-Western approaches to understanding data and digital inequality issues. In many countries, the deployment or diffusion of automated technologies is taking place at a different pace and in distinctive political economic contexts.

With regard to self-regulation, participants debated to what extent market self-regulation would suffice as a means of correcting incidents of data-driven discrimination. In this model, poor, unpopular or unfit-for-practice predictive systems would be forced out of the market 'naturally' through losing customers and potentially going out of business. Others pointed out that our understanding of market functioning and self-regulation mechanisms are based on a faulty assumption of perfect information: in practice, the business of predictive analytics runs on information asymmetries (between user and supplier of services). The absence of perfect information thus makes self-regulation impossible.

Despite these differences of opinion, participants agreed that specific mechanisms in certain 'sensitive' areas of life (e.g. healthcare, finance, personal banking and insurance) should be implemented in order to protect individuals from harm. Examples of such mechanisms can already be found in existing legal frameworks, such as the US health insurance system that now prohibits discrimination based on an individual's pre-existing medical condition(s).

Uncertainty as Technical Solution

With respect to technical solutions, participants examined human-induced bias in both technology design, including specifically the inability of some designers to recognise their own biases. Despite the problem of bias, some participants suggested that algorithmic fairness is nonetheless possible. The predictive analytics that power big data systems thrive on the maximum reduction in uncertainty. The more information or data that is incorporated into the analysis (including indicators that can cause potential results to be biased), the more precise the automated decision or prediction will be. The participants debated whether the introduction of uncertainty in data processing can make predictive analytics more equitable. Needless to say, technical solutions to data-driven discrimination remain works-in-progress, both at the level of culture, such as nurturing a generation of ethical machine learning programmers, as well as computation, such as generating new mathematical designs for fairness.

A Legal Framework that Bites?

The participants turned their attention to reviewing some of the existing legal frameworks, with a particular focus on Europe. They discussed how, at present, the

General Data Protection Regulation (GDPR), which promotes data protection and protects against discrimination and human rights violations, is not fully functional. For example, employers are already prohibited from using data about an individual's socio-economic attributes when making hiring decisions. Yet in practice, employers have other indirect ways through which they can include consideration of these factors in the decision-making process – for example, by looking at a candidate's geolocation and making socio-demographic assumptions on that basis.

The GDPR framework cannot at present address non-transparent use of data, nor can it address cases where data usage is 'unplanned' (i.e. data was not originally produced for the purpose that causes an issue). In short, participants felt that the GDPR does not have a strong mandate and is being inadequately enforced. Participants discussed possible mechanisms to ensure the protection of the most vulnerable people in society, and debated how legal systems should be designed to meet high standards for fairness.

Concluding Thoughts

Throughout the workshop the participants debated issues related to predictive analytics and called for closer examination of issues of fairness, justice and equality. During the workshop, the discussion shifted from a conversation around procedural ideas for combatting data-driven discrimination to one that revolved around normative ideals of justice. The workshop generated a range of questions about individuality, collectivity, equal distribution of data-related benefits, discrimination and market-failure, as well as the future of predictive analytics. These important questions, which have will have such a profound impact on all of our lives, remain to be debated in academic, professional and public discourses.