DATA-POP ALLIANCE
WORKING PAPER

Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics

DRAFT V1 FOR DISCUSSION

December 2015

HARVARD HUMANITARIAN INITIATIVE

mit media lab

DATA-POP ALLIANCE
WORKING PAPER

Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics[1]

DRAFT V1 FOR DISCUSSION

December 2015

Emmanuel Letouzé[2]
David Sangokoya[3]

[1] Prepared for and presented at the Data Associations in Global Law & Policy: A Workshop for New Writing held on Friday 11 December – Saturday 12 December, 2015 at UNSW Australia, Sydney. The authors are grateful to the conference's organizers Fleur Jones, Lyria Bennett Moses and Daniel Joyce. The authors have benefitted from comments from Alex 'Sandy' Pentland during the development of this draft and inputs from Data-Pop Alliance's team (Natalie Shoup, Gabriel Pestre and Lauren Barrett).
[2] Director and Co-Founder, Data-Pop Alliance, Visiting Scholar, MIT Media Lab; corresponding author: eletouze@datapopalliance.org
[3] Research Specialist, Data-Pop Alliance

Table of Contents

You do not want to be an edge case in this future we are building.

Hilary Mason, Fast Forward Labs[1]

Introduction

Algorithms have bad press. They are everywhere yet poorly understood by the general public, referred to as 'black boxes' concealing sophisticated and insidious mechanisms that crunch citizen-consumers' data to make predictions that turn into prescriptions and lock these subjects into their condition. Fan of "Game of Thrones"? Forget "War and Peace".

There is partial truth—yet many caveats and complexities conveniently left out—in this broad summary of the 'risks' posed by the growing reliance on algorithms in various aspects and activities of our lives, chiefly those made 'on our behalf' by corporations and governments. But—as this paper argues, taking a voluntarily optimistic and forceful stance—the rise of algorithms also provides one of those 'historical opportunities', both a practical way and moral obligation, to reengineer current power structures and decision-making processes within data-infused societies in positive ways.

For starters, it is not as if the world nearing the end of 2015 were an earthly democratic paradise threatened only by the ominous prospects of an 'algorithmic future' designed and implemented by machines—computers and robots. Our world faces threats to security, injustice and instability due to political unrest, armed conflict, and climate change. Further, this world has long been governed by algorithms in the form of rules and procedures. Most of these predated the digital and data revolutions and are thus analog; quintessentially human and thus fallible, and more often than not in the interest of the powerful. Of course, new 'algorithms'—new rules—have historically been introduced to improve the human condition, usually following great struggles: think of the civil rights legislation in the U.S. Of course, today we think and talk about different kinds of algorithms that feed on different inputs, seem less 'human', and have more immediate and possibly more powerful effects.

The world is seemingly facing a 'Big Data revolution', as its 193 'leaders'—more or less democratically elected heads of states or governments alongside true autocrats—have committed to supporting the Sustainable Development Goals (SDGs), a set of 17 socio-economic, political and environmental objectives forming and structuring the development agenda of the next 15 years. Against this background, a question one may ask is: what role can algorithms play to make this world a better place? Can algorithms of the Big Data era, and the opportunities, risks and questions they raise, be leveraged as forces of positive disruption?

At the risk of sounding naïve (or hypocritical), we will attempt to argue that the answer is yes. Leveraging algorithms both as practical tools and conceptual levers—as forces of positive disruption—is no small task. But it doesn't seem out of reach. Priorities, requirements and pathways start to be visible. Big Data could be used to grease the gears of democracy by giving greater control to people over the use of the most valuable resource of the 21st century: their data.

The main objective of this paper is to discuss whether and how the future of algorithms, or the future with algorithms, can be crafted such that their development and deployment—from their design to their use, including control, evaluation, auditing, governance—be based on and foster core democratic values such as accountability, transparency, participation, and collaboration. In doing so, we will focus on algorithms affecting public life and policies to maximize benefit for citizens, or 'public good algorithms', but the discussion aims to have broader applicability.

One of the arguments or observations it makes is that in discussions about the implications of Big Data for societies, algorithms have received both too much attention and too little consideration. Too much attention because there has been highly targeted media and public backlash on the nuts and bolds of algorithms as 'black boxes' that needed to be opened, at the expense of richer debates about the purpose of analysis and more importantly the nature of the data being used; too little consideration because algorithms seem to be too swiftly labeled as bad without a thorough enough reflection on the many unique levers and entry points they may offer to serve humanistic principles as part of new data lifecycles and ecosystems.

The rest of this paper is structured as follows. Section 1 attempts to clarify the context and concepts that frame our discussion; section 2 presents use cases of 'public good algorithms', highlighting their positive and negative sides and effects. Section 3 proposes a series of reflections and suggestions for 'using' algorithms as a means and as catalysts to revive democratic ideals. Section 4 discusses the case of a specific public industry-official statistics partnership—and initiative in the making—the Open Algorithms for National Statistics Offices project (OANSO)—to shed additional light on the possibilities and requirements ahead. Section 5 offers concluding thoughts on the possibilities of algorithms for democratic societies.

I. Algorithms in the Big Data Era: Context and Concepts

A. Big data and Big Data: Features and Functions

The Big Data revolution has been described as a "flood"[2], a "tsunami"[3], and a "tidal wave"[4] of opportunities and challenges for institutions and individuals to act upon the analysis of the petabytes of digital signals and traces of human actions and interactions. All of these analogies conjure the accurate sense that the amount of data collected is greater than ever before. Our capacities to understand these data have advanced greatly in the recent past and continue to do so. Through conferences, working groups and academic networks, the communities considering the implications surrounding this data have also grown.

While it was once framed as the "3 V's" of big data (volume, velocity and variety) around 2008-12, we prefer to conceptualize Big Data as an ecosystem made up of "3 C's", introduced in earlier papers[5]:

1. The C of *crumbs*—i.e. those *"digital bread crumbs"* or those "digital translations of human actions and interactions passively emitted and captured by digital devices". At the center of our information societies is the production of massive amounts of data through connected platforms, social networks, and machines. This feature is important as it presides over a fundamental qualitative shift as much as a quantitative one and gives Big Data its deeply political nature.

2. The C of *capacities;* i.e. tools and methods to collect, aggregate and analyze data. Algorithms—to be defined and discussed below—fall squarely under capacities, and stand firmly at the center of this ecosystem, as both products and drivers of its expansion. Parallel computing is another key aspect without which Big Data would not exist as a techno-social phenomenon as it allows making computations in a fraction of the time—sometimes years—it would take to run them on one machine.

3. The C of *communities*—i.e. all those involved in generating, governing and using data, including data producers, end users, policymakers, experts, privacy advocates and civic hacker communities. Namely, groups. To date the two constituencies that have been the most active in levering algorithms to make decisions of not as the centerpiece of their business are large private companies and government agencies—notably those in charge of surveillance activities—with academia coming third and organized advocacy groups and networks (e.g. in the humanitarian space) coming fourth.

A number of additional conceptual notes are called for on the features of big data and the functions of Big Data. First, the 'big data' on which algorithms run come in two main forms, with important implications. One is unstructured data – such as videos, photos, tweets, etc. In almost all cases, these are produced for some other purpose than the one they are subjected to through automated analysis. Unstructured means the information the dataset contains and displays is not provided in a structured format—rows and columns.

Part of the automated analysis process will involve a structuration or 'cleaning' step. Some have called these unstructured data "signals" as opposed to "traces". Their metadata (e.g. where the photo was taken) are structured. Another kind is indeed structured data—or "traces". Fundamentally, structured data, of which Call Detail Records have become the most cited examples, or credit card transactions, are *answers* to questions asked by the data collector: where, how many, how long, etc. The results are placed in columns and rows.

The distinction matters for two main reasons. One, algorithmic analysis 'works' typically best—under some estimation metrics to be discussed—when combining both kinds[6]; second and relatedly because the collectors of the structured data have much greater control over the data and are able to restrict access to them both legally and technologically.[7] Both kinds raise complex and highly debated ethical and political questions about collection, storage, use and control modalities that are at the heart of the issue being discussed.

Second, in this paper, we also refer to the four-tier taxonomy of 'functions' of Big Data developed in previous papers, all of which can and generally involve algorithms; namely:

1. *Descriptive,* which involves narrative or depiction of some human phenomenon thorough maps word clouds or visualization;
2. *Predictive*, which includes what has been called 'now-casting'—i.e. making real-time inferences on some phenomenon based on cell-phone activity for instance–as well as forecasting what may happen next, both based on past observed trends and patterns,
3. *Prescriptive*, understood as going beyond description and inferences to establish and make recommendations on the basis of *causal* relations, for instance by identifying the effects of a public transportation system on public safety;
4. *Discursive,* which concerns spurring and shaping dialogue within and between the 'Big Data communities' about policy efficiency, data use, and, as a case in point, algorithmic use.

Evidently, one of the main characteristics or consequences of traditional algorithmic analysis—and somewhat of a semantic challenge—is the blurring of lines between prediction and prescription; if Amazon's algorithm predicts that Justin may like that pop album (because many other Justin's do), it will 'prescribe' (or suggest) it to Justin. We discuss later the limits of this and potential ways around it.

B. *The appeal, functions, uses and risks of algorithms*

Algorithms are a logical series of steps help us find answers and generate value amidst the chaos of data; more advanced algorithms (through machine learning, as will be discussed later) can adapt and, based on previous observations, make predictions and recommendations. a set of "encoded procedures" or "a logical series of steps for organizing and acting on a body of data to quickly achieve a desired outcome."[8]

Based on their function and power, Diakopoulos (2014) characterizes the taxonomy of algorithms in four broad categories[9]:

1. **Classification**: categorizing information based on its features into separate classes;
2. **Prioritization**: associating rank and emphasis on particular information or results at the expense of others through a set of pre-defined criteria (e.g. search engines, efficiency management algorithms, etc.) – process decision with respect to what you're doing
3. **Association:** determining correlated relationships between particular entities via semantic and connotative abilities;
4. **Filtering**: including and/or excluding information as a result of a set of criteria

The table below provides examples of types of algorithms across these categories.

Table 1. – adapted from Diakopoulos (2014) and Latzer et al (2015).

Function	Type	Examples
Prioritization	General search engines	Google, Bing, Baidu
	Special search engines	Genealogy; image search; Shutterstock; social media Info.com
	Meta search engines	Yummly
	Semantic search engines	Quora, Ask.com
	Questions & answers services	
Classification	Reputation systems	Ebay, Uber, Airbnb
	News scoring	Reddit, Digg
	Credit scoring	CreditKarma
	Social scoring	Klout
Association	Predictive policing	PredPol
	Predicting developments and trends	ScoreAhit, Music Xray, Google Flu Trends
Filtering	Spam filter	Norton
	Child protection filter	Net Nanny
	Recommender systems	Spotify, Netflix
	News aggregators	Facebook News Feed

The shift to utilizing algorithms (or rather, systems of algorithms) for decision support is not very surprising given incentives towards data-driven decision-making for both companies and governments—which some have described as an effect of the rise of neoliberal policies in the 1980s.[10] Data provides structure for classifying and understanding phenomena and when collected, aggregated, and analyzed provides capabilities for drawing out unexpected insights, trends or predictions. With a flood of data to mine and the ability to make statistical predictions and recommendations, it's again no surprise that companies and public sector actors are turning to algorithms to solve complex problems at the limits of human decision-making. Taking a historical perspective, the history of human decision-making—particularly in positions of power over determining resource allocation, fairness, justice, and other public goods—is wrought with innumerable examples of extreme bias leading to inefficient and / or unjust processes and outcomes. Consider slavery, global histories of discrimination, or gender inequality. In short, human decision-making has been significantly limited and the turn towards algorithms represents the hopeful search for objectivity, evidence-based decision-making, and a better understanding of our resources and ourselves.

II. Cases, Benefits and Harms of Public Goods Algorithms

This paper places emphasis on what we call public goods algorithms—algorithms strongly influencing decision-making and resource optimization for public goods such as justice, public safety, health, access to finance and fair employment. These algorithms are particularly consequential given the magnitude of their impact, effect on essential quality of life, information asymmetry surrounding their governance, and the social disenfranchisements they can reinforce or create themselves.

A. Expected value and actual use of public goods algorithms

Gillespie (2013) described 6 dimensions of these algorithms, including[11]:

1. patterns of inclusion—i.e. the "algorithmic" process determining data selection and preparation;
2. cycles of anticipation—i.e. "consequences of attempts by those creating the algorithms to have informed on users and make predictions");
3. evaluation of relevance—i.e. filtering criteria for algorithms to determine relevance and legitimacy);
4. promise of objectivity (presentation of impartiality);
5. entanglement with practices—i.e. other "algorithmic" processes in which users change their behavior to fit algorithms; and the
6. production of "calculated publics"—i.e. "process of algorithmic presentation of publics back to themselves and how this shapes a public's sense of itself").

These algorithms are not just consequential in and of themselves, but also given their interactions with other seemingly disconnected, linked data and algorithms, the unveil more variable relationships related to people, their networks, and their resources. Resource allocation optimization forms a strong component of the rationale for using public goods algorithms—the need for precise decision-making over limited resources. Decision optimization in this regard is not new.

However, the level at which we can assess vast amounts of personal data/bread crumbs, the capacities that can quickly analyze and deliver results, and the communities of experts and common people who hold these results to be objective in some way creates a new kind of decision optimization facilitated by algorithms and the data from which they are based.

Table 2. Public goods algorithms

Public good	Example	Function(s)	Purpose
Justice	Recidivism models for sentencing and parole	Classification Association	To assess possibility and associated risks of recidivism relative to justice
Public safety	Predictive policing	Prioritization Association	To locate areas of ongoing and burgeoning criminal activity in a given place
Fair access to finance	Credit and loan access	Classification Association	To provide fair access to finance options, particularly for vulnerable populations
Fair employment	Employee hiring	Prioritization Filtering	To promote fair employment process and prevent discriminatory bias

In each of the case studies below, we assess the following: the nature of the public good algorithm; function and purpose; governance structure and decision makers; the nature of the legal/regulatory/accountability structures around them.

Box 1.1: Case study: Predictive policing

Predictive policing, or a "formal, quantitative research program"[12] as some suggest, is when police employ mathematical, process-oriented data in order to anticipate a crime, with the goal of preventing it from happening and/or mitigating its effects. This is used in conjunction with and complements traditional policing, in which non-quantified and thus non-algorithmic data is often gathered to anticipate crimes.[13] As a public good, predictive policing could generate value by reducing excess human capital (resource maximization), increasing response rate to crimes already committed, reducing historical-based profiling (algorithmic profiling remains), and reducing overall crime rates.

There are four ways to utilize predictive policing:

1. Methods for predicting crimes: These are approaches used to forecast places and times with an increased risk of crime.

2. Methods for predicting offenders: These approaches identify individuals at risk of offending in the future.

3. Methods for predicting perpetrators' identities: These techniques are used to create profiles that accurately match likely offenders with specific past crimes.

4. Methods for predicting victims of crimes: Similar to those methods that focus on

offenders, crime locations/hotspots (or crime mapping[14]), and times of heightened risk, these approaches are used to identify groups or, in some cases, individuals who are likely to become victims of crime.[15]

Algorithms are alive and thus inherently are changing constantly if we allow them to. They have the capacity to change more rapidly than laws do, are more agile, more disruptive, and could even potentially shed light on a new form of representative democracy. Thus, governance (here referring to human capital) plays three major roles in predictive policing: 1) to ensure a paucity of data gaps, that otherwise would lead to false positives in predictive policing; 2) to vet accuracy of data entered (and cleaned and stored); and 3) to audit the precision of algorithmic functions.

With predictive policing, it is crucial to ensure that the algorithmic outputs are not used as *prescriptive* material - that is, suggesting a person may commit a crime when they have not, and/or will not. An example of this is (most often, racial) profiling - police are up to 28 times more likely to use stop and search procedures on black people versus white people, and 25% more likely to include an advertisement suggesting arrest when a black name is searched on Google.[16] The former is an example of clinical (human-based predictive) profiling; the latter is an example of algorithmic profiling - which could be a harmful byproduct of predictive policing. As for the former, with predictive policing we could move from clinical to statistical processes.[17] Perhaps in the future this would mean figures for being stopped and searched would start to match the low 3% arrest rates of all of those who were stopped and searched.[18]

Box 1.2: Case study: Recidivism models for sentencing and parole

In the United States, parole boards are tasked with periodically reviewing inmates' cases to determine whether or not to release them. One of the major challenges they face is balancing the interests of the prisoner (not being kept in jail if she/he is deemed fit to reintegrate society) with the interests of the rest of society (not being exposed to the additional risk of the release of a prisoner who was unfit for release). Many factors come into play in assessing whether or not the release of a prisoner will be beneficial to both that individual and the rest of society: ability to find a job, presence of a support network to return to upon release, likelihood of committing another crime, etc.

This last point, the issue of recidivism, has long been one of the major focuses of parole boards in the US. Traditionally, qualitative interviews were used to assess whether or not an individual was fit for release, and "factors like the severity of a crime or whether an offender shows remorse"[19] were weighted heavily in parole rulings.

Since the 1920s, much focus has been put using more quantitative approaches to forecast whether or not a given individual would commit another crime within a certain number of years of their release. These forecasts are based on "a mixture of factors such as age, race, prior offense history and school grades to determine whether an inmate should be paroled or not"[20].

This type of forecasting typically involves algorithms that assign weights to various factors such as type of crime, age when the crime was committed, number of previous convictions, current age, marital status, level of education attained, etc. "Some assessments analyze as many as 100 factors, including whether the offender is married, the age of first arrest and whether he believes his conviction is unfair. In Texas, a rudimentary risk-assessment measures just 10 factors."[21] This information is generally gathered from the inmates records or collected during an interview with the individual.

Each question is given a weight and each possible response to a question is given and number of points (either positive or negative), based on how indicative they are of the possibility of a repeat offence. The results of each section are compiled and, based on the final score, the prisoner is assigned a category of how likely it is that she/he will commit a new crime within a set number of years, typically three, from her/his released. Such forecasting mechanisms differ widely in their implementation, but the general principle is that responses to the questionnaire have fixed weights and point values (which may differ from state to state).

In some cases, these classification algorithms are preceded by a filtering algorithm, since inmates must generally meet certain criteria (such as having served a minimum percentage of their sentence and having at most a certain number of [[[misbehaviour complaints]]] on their record).

Some states use the category (ex: "low", "moderate", "high") of individuals' risk of recidivism as an in the decision of whether or not to release them, with the final decision coming from a judge or review board; others convert the results directly into a decision by choosing to release all prisoners that are below a certain risk threshold. Such thresholds may differ from place to place: states that face overcrowding in their prisons may well be willing to accept the political and societal risks of release higher numbers of prisoners each year. Texas, for instance, under its current policies, releases about 40% of inmates who are found to be moderately likely of committing another offense.[22]

These algorithms may be implemented manually (by having an interviewer compile the results in a questionnaire and calculate the scores) or electronically (by entering the data into a computer system which calculates and returns the result). Electronic implementations include, for example, Compas, a system designed by Northpointe Inc., which is currently in use in Michigan, where computerized assessments were introduced in 2006 and adopted statewide in 2008[23]. Such assessment systems often rely heavily on research by criminologists, who help determine which factors should be considered for making accurate and ethical forecasts.

In addition to the classification and filtering algorithms described above, associative algorithms may also be used in the design of the categorization algorithms. For instance, historical data on prisoner releases and cases of recidivism can be processed with machine learning algorithms to determine which factors are most often associated with repeat offenses. These results are used to help determine the weights and scores used in the questionnaires, which are also reviewed -- and often designed -- by criminologists[24].

Although the use of historical data can lead to accurate predictions of how likely someone is to commit another crime, certain models[25] omit this type of data because of the pitfalls it presents: people of certain races, demographics, and groups may be affected disproportionately by the trends that emerge from historical data, either because of omitted variable bias, or because the historical records themselves capture discriminatory or inefficient practices in law enforcement, prosecution, or sentencing that were in effect when the data were collected. Similar caveats also apply to using data to validate the efficiency of such predictive models: "experts say it is difficult to measure the direct impact of risk prediction because states have also taken other steps to rein in corrections costs, such as reducing penalties for drug offenses and transferring inmates to local jails."[26]

These processes, therefore, should be scrutinized, for the same reasons decision optimization has always been challenging: the biases of resource managers and, in the public sector, how these biases affect minority groups — both visible and invisible. By visible, we describe the defined legal categories in existence today (such as race, gender, sexual orientation, etc). As Barocas and Selbst (2015) and many others have described in the literature, there is clear evidence that data we feed algorithms are inherently biased—in addition to the coded variables generated by machine learning—[27] and can lead to harm for visible minorities and invisible minorities—groups of unknown users with similarities related to the rules of the algorithm or undisclosed/previously undetectable patterns (e.g. users without middle names in an algorithmically-run system that only reads names with middle names).[28]

B. *Nature and magnitude of algorithmic harms*

Many authors and experts have underlined the risks of algorithmic harms through examples such as the 2014 Facebook Newsfeed emotion contagion experiment,[29] the 2012 plans of a German credit agency to mine Facebook and other social media data to assess creditworthiness,[30] and models to predict potential recidivism in criminal offenders (known as recidivism models).[31] The use of personal data in such algorithms (often without the knowledge of users) introduces a tension between efficiency and risks to privacy and the inability to opt out. Through expert discussions and events such as NYU's Governing Algorithms and Accountability conferences, dialogues have become widespread on the legal, ethical and normative frameworks governing algorithms and their associated harms, risks and impact for decision support.

Mason's tweet on edge cases offers a blunt realization to a grave concern the ongoing, unexplored, dangerous use of algorithms for decision support: a digitized, justified, "objective" perpetuation of existing inequalities in today's society. Whether as visible, legal minorities (such as by race, gender, sexual orientation, etc.) or invisible minorities (as minority groups of users not considered by algorithms), being an edge case means being systematically disenfranchised without opportunity for redress either in the eyes of the algorithm or those with a particular set of skills to even begin understanding the algorithm's newly established rules of the game.

Uncovering the nature of these systems of algorithms—particularly algorithms influencing public resource optimization—leads to a series of questions related to algorithmic harms and accountability: What are the implications as public sector actors use algorithms for decision support for using recidivism models for sentencing and parole or predictive policing? Who decides the nature of level of sufficient accountability for users, particular edge cases? What mechanisms can society use to address the accountability concerns regarding algorithmic use? Ultimately, are there existing conditions and opportunities for algorithms to be used to positively disrupt the limits of human decision-making and if so, what would be the requirements and elements making and shaping such a disruption?

An immediate tension here is in the use of personal data, particularly without the consent of the consumer. A lack of transparency here invites the notion of algorithm as gatekeeper (whether to give implicit utility to the user or silently manipulate). The user remains in the dark until evidence of manipulation is made public (this can occur more readily for some algorithms than others) and in these instances, it's not clear either how redress wrongs or prevent further ailments.

Another tension lies in the promise of transparency; on the one hand, transparency is often an accountability measure to give agency for users to recognize and understand wrongs. However, users will not necessarily be able to identify wrongs by making the algorithm's "rules of the game" transparent, nor even observing the semblance of such instances change the nature of the biases inherent in the data.

As more of our tasks and decisions are delegated to and through algorithmic procedures, tensions in addressing informed consent and redressing potential harms resulting from algorithms warrant need for feedback or accountability measures to check their use and power.

Several researchers and experts have contributed to a growing literature on assessing the opportunities and challenges associated with various algorithmic governance frameworks, particularly advocating for an overall need for algorithmic transparency. Latzer et. al (2014) provide a comprehensive assessment on algorithmic governance highlights a range of options and limitations "on the continuum between market solutions and state regulation" assessed across nine areas of algorithmic risk.[32] [33] Adapted from Latzer et al.'s (2015) assessment, the table below highlights the range of options and limitations in addressing algorithmic risks and harms available to the state and individual companies, as well as within industries and market economies.

Table 3. Range of options and limitations of algorithmic governance frameworks

	Options		Limitations
Market solutions: Demand side	Consumer-driven solutions for data protection and privacy	Tor, virtual private networks (VPN), OpenDNS, privacy-enhancing technologies (PETs)	No available opt-out options from current providers, information asymmetry, data and computational literacy
Market solutions: Supply side	Product innovations	New/modified services reducing risks via business strategies (e.g. nachirchten.de, search engine DuckDuckGo)	Widespread consumer use and dominance of traditional services (Google, Facebook)
Companies: self-organization	Company principles and standards Internal quality	E.g. search neutrality, minimum principle of data collection[34]	Lack of incentives for data protection standards

	assessments, complaint feedback Internal ethical boards for user experience	Google's ethics board, in-house algorithmists[35]	"Transparency dilemma"[36] Limited options for B2B markets
Industry/Branch: self-regulation	Codes of conduct Organizational and technical industry standards Quality seals and certification bodies Ombudsmen and arbitration/mediation boards Ethics committees	E.g. Digital Advertising Alliances in Europe/USA, Creative commons licensing, digital rights management (DRM), do not track	
State intervention	Regulation	Microsoft, Google Post-Safe Harbor	

III. Conditions and contours for spurring positive disruption by leveraging algorithms

A. Merits and limits of algorithmic transparency and accountability

As noted in the previous section, both governments and companies use several kinds of algorithms for decision support and optimization. For public good algorithms, accountability in government and corporate use of such powerful decision support tools is fundamental in both validating their utility toward the public interest as well as redressing corrupt or unjust harms generated by these algorithms.

The nature of various public goods algorithmic use, as well as the lack of computational literacy among citizens, makes algorithmic transparency difficult to generalize and accountability difficult to assess; what mechanisms lie in both the public and corporate domain for addressing consent and discrimination challenges associated with algorithmic use? To what extent do these mechanisms allow for citizens to negotiate the use of their data and additionally ask questions of the black boxes of information providing critical decisions on central facets of public life? In this section we explore the applications, implications and challenges of existing algorithmic accountability mechanisms in addressing lack of consent and discrimination issues.[37]

However, what still entrenches all of these options is constant interplay between a history of powerful actors, poor quality data, and historically-bound algorithms.

B. Requirements for positive disruption of public goods algorithms

As noted earlier several scholars have emphasized the detrimental effects of algorithms using historical data—chief of which their self-fulfilling effect. Despite existing strategies for engineering accountability mechanisms to address algorithmic harms, algorithms using historical data will be intrinsically bound and entrenched by the biases both in design and further continued via machine learning. These arguments typically call into question the use of machines in decision support and the need to protect the role of human decision-making. However, are algorithms therefore unable to positively influence and support human decision-making, or perhaps are there still opportunities in using these tools for decision support?

Two requirements towards what we frame as a "positive disruption" of algorithmic decision-making for public good involve reconsidering the impact of existing human algorithms and subjectivity, and renegotiating a new deal on data stewardship between citizens and their data.

Across the case study examples of recidivism models and predictive policing, human algorithms—existing logical series of rules and processes leading to the generation of specific outputs—form the bases of these activities and algorithms used. Sometimes these human algorithms are quantitative (i.e. processes governing development of checklists used toward recidivism model generation) or qualitative in nature (i.e. expert opinion governing where police go). Regardless, these processes are driven toward optimizing for a binary determination that we assume must exist (with a margin of error of zero): criminal/not criminal; good loan candidate/bad loan candidate; valuable employee/not valuable employee.

Towards such ends we inherently (and irrationally) associate weights to numerous factors in order to classify, prioritize and filter information in order to arrive at this binary decision. Furthermore, other assumptions in these activities bias both process and data selection efforts toward specific solutions or theories: efforts to protect public safety focused on a policing solution or the political costs vs. social impact of one public case of recidivism for other prisoners to be paroled. This can be particularly true within the political environment of public resource allocation.

In short, human algorithms dictating the assumptions and processes surrounding data selection and collection form the basis of our technical, machine-based algorithms. The implication here is that machine-based algorithms used for decision-making are objective relative to the input data and its given assumptions generated by human decision processes. It may then be no surprise that the machine-based algorithms based on these processes both discriminate and fail to produce output at a margin of error in the same way as humans; this is not as a result of being machines but as a function of their input and perhaps the failed hypothesis dictating the input process. Reconsidering the impact and biases coded by human decision processes into the data invites the designer to revisit missing gaps in data selection, re-weight particular variables, and use the algorithm as a tool for testing and considering potential approaches in addressing public resource optimization (for example, the location of "crime hotspots" via CDR data as an opportunity to explore non-policing, environment-focused approaches to crime).

Secondly, the other component in addressing algorithmic accountability lies in resolving the evolving issue of consent in the digital age. Currently, Internet and mobile users are at the mercy of a litany of out-dated privacy policies, complicated service agreements with Internet service providers, and the ability for local data protection authorities to enforce the law in an ever-changing technological and digital environment. For privacy rights to become more meaningful, they must also be exercised more meaningfully by the data subjects themselves.

It is in this view that Greenwood, Pentland, et al have fleshed out a "New Deal on Data"—a renegotiation of the relationships between citizens and their own data that involves gaining "key rights over data that are about them."[38] Their suggestion draws inspiration from the E.U.'s data protection directive, which has, since 1995, successfully altered the practices of major service providers.[39] The authors suggest going beyond such laws, enshrining a system of specific individual control over each piece of personal data. In order to achieve this, the authors envision a "trust network" enabled by the alliance of law and technology: on the technological side, all items of data can have "attached labels specifying where the data came from and what they can and cannot be used for."[40] The terms on the labels can, in turn, be matched by the terms of art used in the legal system (in contracts, regulations, etc.) An efficient network would require international harmonization in order for the various labels to be compatible amongst each other and for legal terms to be translated without loss of meaning. While this might seem complex, Greenwood et al underline that it is akin to Visa Operating Rules and, more generally, the way the credit card network operates.

Their proposal flows naturally from existing systems of privacy protection through user consent, and from the desire to make such consent more fine-grained, more informed, and more genuinely free. While Greenwood et al's approach presents a pragmatic, practical way to hand over data control to users, it comes with new risks of its own. The authors view personal data as a "new asset class," and, in order to encourage positive uses of this data, suggest both "viewing data as money" and creating incentives to share it.

While this solves the problem of having massive caches of information "siloed" within private companies, cordoned off from many potentially beneficial uses, it also risks increasing inequalities. If services or financial incentives are provided in exchange for personal data, it is easy to imagine that the most vulnerable populations in society will be more eager to part with their information, thereby making privacy both a privilege and a luxury good. This can be contested, at least in part, by improving data literacy and privacy awareness.

Elements of positive disruption

With the previous requirements forming a basis, a focus on positive disruptions of public good algorithmic use moves away from whether algorithms are good or bad, or whether machines can fully or incompletely express the range of human decision-making; positive disruptions involve elements that reinforce the use of algorithms as tools generating value while safeguarding minorities and edge cases from the realities of human biases.

Features and conditions of these disruptions include:

- Greater **participation** of audiences for scrutiny and meaningful use;
- Greater access to **information** on input data and algorithms; and
- Further opportunities for **education** on data and computational literacy and other demand-driven efforts decreasing potential harms

IV. The 'Open Algorithms for National Statistical Offices project' as a tool for positive disruption

A. Rationale and Genesis

One example of a project that reflects and may help foster these broad principles and objectives is the 'Open Algorithms for National Statistical Offices project' currently being developed by a group involving Data-Pop Alliance. Open Tracking progress of the newly adopted Sustainable Development Goals (SDGs) will increase the demands on constrained and financially pressed National Statistical Offices (NSOs) to collect and analyze data in new areas. In fact, knowledge and data gaps are among the biggest challenges that the development community is facing, according to the UN Report "A World that Counts" launched in 2014. The rapid spread of new technological devices in the developing world has created new vast amounts of data flows that hold great potential if they are processed and analysed. In addition, new and promising sources of data have emerged in recent years, e.g; use of Call Data Records (CDR) to produce proxies of population densities, poverty index, literacy rates, etc. NSOs, among other actors, have expressed a strong interest in accessing this new world of data in a responsible manner.

The analysis of big data is now commonplace for private companies and in fact, most of what we consider big data is collected by them. But it has also caught the attention of public bodies that are starting to understand the potential impact of big data and new technologies. As the UN Report lays out, public health researchers are gaining valuable insights from using anonymised mobile phone data on human migration and linking this to the spread of malaria and dengue fever.

The objective of OANSO (Open Algorithms for NSOs) is to help NSOs access to and/or collect better data and create new ways of understanding it, by developing an Open Software Platform and Open source Algorithms that would enable them to collect indicators based on data collected by private companies (ex: mobile phones or banks) in developing countries, in a safe, private and secure way.

This project is also expected to favor interactions and collaborations between NSOs that have for a long time dismissed Big Data as hype; the private sector that in some cases has dismissed NSOs as outdated, slow, etc; and citizens-consumers that are the primary emitters of the data used.

B. Modalities and expected effects

The structure of the project involves developing a version of the platform by creating a set of services covering the basic needs for data collection, processing, security and audit mechanism, and based on mobile data. This software will be installed and used within a Telco's premises, working on anonymous extract of CDRs. This will mean that no data are being copied and send around, making it safe for competition and privacy as only pre-agreed aggregated results of algorithms are exported. The principles of Open Algorithms and its architecture should be developed to be re-usable later without too much effort for accessing other types of data from banking or retail to create additional indicators. Other tasks include developing the Algorithm certification and usages audit mechanisms (ensuring that only the right version of the agreed algorithms can run on the platform) and country selection.

The project will involve citizen groups who will have a say in deciding which indicators they deem worth investigating, as well as be consulted and trained on the algorithms. This project although currently only in the design phase, could pave the way for the development of other initiatives where 'algorithms' are participatory and open by design.

V. Concluding thoughts

There are many good reasons to fight a 'blind algorithmic future' where those statistically more prone to commit a crime would end up jailed. As noted by Christian Lous Lange: "Technology is a good servant but a dangerous master." The path ahead involves crafting a future where algorithms serve democratic principles and objectives—instrumentally and intrinsically.

Algorithms are already able to help guide public interventions but also individual decisions in a wide range of situations. Systems that rely solely on human intuition or predispositions are far from perfect and human interventions often means that prejudice rather that the law rules. A famous study showed that sentences handed out by judges differed widely for similar offenses depending on whether they were pronounced before or after lunch. Algorithms don't get hungry, nor angry. How many people have been sentenced to death because of a combination of racial biases and bad data? The outcomes of an algorithmic procedure can be evaluated in light of its stated objectives, and eventually adapted. It doesn't mean the end of human intervention; rather it demands human oversight.

Intrinsically, the rise of algorithms may provide an opportunity to reshape power and incentive structures. With corporations and governments increasingly eager to rely on data, data emitters and subjects should be incentivized to demand greater direct control over their data—to weigh in on how, by whom and for what purpose it is used. Technological solutions are in sight—such as OpenPDS for example.

Whereas the prospects of a future where citizens are able to leverage their data and how they are used in ways that foster democracy may seem utopian, it is worth reminding ourselves that the first 'data generation' of humans is 5 about years old. The rate of socio-technical innovation and change is unlikely to abate; there is nothing deterministic about where it will take us. When this generation starts having children, around 2030, will it really be a terrible outcome that more of them grow up in societies using algorithms than ruled by autocrats?

1 Mason, H. [hmason]. (28 October 2015). You do not want to be an edge case in this future we are building [Tweet]. Retrieved from https://twitter.com/hmason/status/659412072720044032

2 Melamed, C. (2014). The data revolution is coming and it will unlock the corridors of power. The Guardian. Retrieved from http://www.theguardian.com/global-development/poverty-matters/2014/oct/01/data-revolution-development-united-nations

3 Forbes,. (2015). *IBMVoice: Four Ways To Innovate Using Big Data And Analytics*. Retrieved 1 December 2015, from http://www.forbes.com/sites/ibm/2014/12/08/four-ways-to-innovate-using-big-data-and-analytics/

4 Bill Franks. 2012. *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics* (1st ed.). Wiley Publishing.

5 Letouzé, E. "Big Data and Development Overview Primer". Data-Pop Alliance, SciDev.Net and the World Bank (2015)

6 Cardon, Dominique. A quoi rêvent les algorthimes. Nos vies à l'heure des *big data*. 2015

7 There are several caveats that are worth making; Facebook data is proprietary even as users may access and analyze photos and posts of their friends (usually); Tweets are public but technologically difficult to harvest at large scale beyond the 'fire hose' and access to historical tweets is expensive.

8 Gillespie, T. (2014). 9 The Relevance of Algorithms. Media technologies: Essays on communication, materiality, and society, 167.

9 Florian Saurwein, Natascha Just, Michael Latzer, (2015). "Governance of algorithms: options and limitations," Info, Vol. 17 Iss: 6, pp.35 – 49.

10 Cardon, Dominique. A quoi rêvent les algorthimes. Nos vies à l'heure des *big data*. 2015

11 Gillespie, T. (2014). 9 The Relevance of Algorithms. Media technologies: Essays on communication, materiality, and society, 167.

12 https://www.palantir.com/2012/09/predictive-policing-a-window-into-future-crimes-or-future-privacy-violations/

13 http://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/welcome.aspx

14 Bogomolov A, Lepri B, Staiano J, Letouze´ E, Oliver N, Pianesi F, Pentland A (2015) Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. Big Data 3:3, 148–158, DOI: 10.1089/big.2014.0054.

15 Perry, Walter L., et al. "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations." RAND Corporation. 2013. http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf

16 http://www.theguardian.com/uk/2012/jun/12/police-stop-and-search-black-people
http://www.huffingtonpost.com/2013/02/05/online-racial-profiling_n_2622556.html

17 https://www.palantir.com/2012/09/predictive-policing-a-window-into-future-crimes-or-future-privacy-violations/

18 http://www.theguardian.com/uk/2012/jun/12/police-stop-and-search-black-people

19 (http://www.wsj.com/articles/SB10001424052702304626104579121251595240852).

20 http://theconversation.com/new-models-to-predict-recidivism-could-provide-better-way-to-deter-repeat-crime-44165

21 http://www.wsj.com/articles/SB10001424052702304626104579121251595240852

22 http://www.wsj.com/articles/SB10001424052702304626104579121251595240852

23 http://www.wsj.com/articles/SB10001424052702304626104579121251595240852

24 http://www.wsj.com/articles/SB10001424052702304626104579121251595240852

25 http://theconversation.com/new-models-to-predict-recidivism-could-provide-better-way-to-deter-repeat-crime-44165

26 http://www.wsj.com/articles/SB10001424052702304626104579121251595240852

27 Barocas, Solon and Selbst, Andrew D., Big Data's Disparate Impact (August 14, 2015). California Law Review, Vol. 104, 2016. http://ssrn.com/abstract=2477899

28 Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*.

29 http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/

30 http://www.spiegel.de/international/germany/german-credit-agency-plans-to-analyze-individual-facebook-pages-a-837539.html

31 http://theconversation.com/new-models-to-predict-recidivism-could-provide-better-way-to-deter-repeat-crime-44165

32 Risk categories include "manipulation, bias, censorship, violation of privacy rights, social discrimination, violation of property rights, abuse of market power, effects on cognitive capabilities and heteronomy."

33 Florian Saurwein, Natascha Just, Michael Latzer, (2015). "Governance of algorithms: options and limitations," Info, Vol. 17 Iss: 6, pp.35 – 49.

34 Introna and Nissenbaum 2000; Langheinrich 2001; Cavoukia 2009

35 Lin and Selinger 2014; Mayer-Schonberger and Cukier 2013

36 Bracha and Pasquale 2008; Granka 2010; Rieder 2005; Latzer et al. 2002, 2007

37 Pasquale, F.: https://aeon.co/essays/judge-jury-and-executioner-the-unaccountable-algorithm
http://theconversation.com/big-data-algorithms-can-discriminate-and-its-not-clear-what-to-do-about-it-45849
Latzer et al – Governance of Algorithms: Options and Limitations - http://www.mediachange.ch/media//pdf/publications/GovernanceOfAlgorithms.pdf

38 Daniel Greenwood, Arkadiusz Stopczynski, Brian Sweatt, Thomas Hardjono, and Alex Pentland, "Institutional Controls : the New Deal on Data"
39 E.g. Google « right to forget » decision of 2014.
40 *Id.*