Algorithmic Regulation

NAS-Royal Society Sackler Forum , The Frontiers of Machine Learning
Washington DC, 31 Jan-2 February 2017

Professor Karen Yeung
Director, Centre for Technology, Ethics, Law & Society (TELOS)
The Dickson Poon School of Law, King's College London

A digital revolution for government?

- A new Industrial Revolution is dawning, made possible by networked digital technologies, cloud computing and powered by the engine of 'Big Data'

- Evidenced by rapid take up in the commercial sector, transforming service delivery across a wide range of industries (incl finance, marketing, retail, media, health, dating etc)

- Why not apply the same techniques used by Google, Amazon, Facebook and their ilk to the task of regulatory governance ?

From critical data studies to 'algorithmic regulation'

- Large and growing literature concerned with 'critical algorithm studies'
- But, yet to be explored through the lens of regulatory governance scholarship framed around the concept of '**algorithmic regulation'**
- Ongoing thoughts – critical feedback welcome

Outline

- Aim: critical examination of ‘algorithmic regulation’ (various lenses from social science and legal scholarship)
- Define algorithmic regulation
- Examine the logic of algorithmic regulation: a proposed taxonomy, with two basic forms (reactive vs predictive)
- Algorithmic regulation as a system of social ordering

a) The politics of algorithmic regulation
b) The new surveillance
c) Algorithmic power, accountability and the social foundations of democracy

- Conclusion

Algorithmic regulation

- O'Reilly (2013) does not define algorithmic regulation, but merely points to various technological systems which he claims share four features:

1. a deep understanding of the desired outcome
2. real-time measurement to determine if that outcome is being achieved
3. algorithms (i.e. a set of rules) that make adjustments based on new data and
4. periodic, deeper analysis of whether the algorithms themselves are correct and performing as expected (O'Reilly 2013).

- Eg. motor vehicle fuel emissions systems, airline automatic pilot systems, credit card fraud detection systems, drug dosage monitoring by medical professionals, internet spam filters and general internet search engines

A working definition

- **Algorithmic regulation** refers to regulatory governance systems that utilise algorithmic decision making

- What is regulation or a 'regulatory governance' system?

- "intentional attempts to manage risk or alter behaviour in order to achieve some pre-specified goal" (Black 2014).

Note:

- regulation is primarily undertaken by governments but also pursued by **non-state actors** and entities (Black 2008).
- size of the regulated 'population' is highly variable. Eg individual fitness tracker vs Uber ride sharing platform.
- regulation is an intentional activity directed at achieving **a pre-specified goal**, so any regulatory system must have some kind of system 'director' (or 'regulator') to determine the overarching goal of the regulatory system

What is meant by ‘algorithmic’?

- Algorithms are encoded procedures for solving a problem by transforming input data into a desired output, based on specified calculations and procedures (Gillespie, 2013) – not necessarily software

- Software engineering perspective: technical understanding of algorithms as referring to the logical series of steps for organising and acting on a body of data to achieve a desired outcome quickly, which comes after the generation of a model – ie the formalisation of a problem and the goal in computational terms

- But for social scientists – core concern is the larger sociotechnical assemblage that includes the algorithm, model, target goal, data, training data, application, hardware – all connected to a broader social endeavour aimed at knowledge production (‘algorithmic system’)

Algorithmic regulation – working definition

Algorithmic regulation refers to decision-making systems that regulate a domain of activity in order to manage risk or alter behaviour through continual *computational* generation of knowledge from data emitted and directly collected (typically in real time on a continuous basis) from numerous dynamic components pertaining to the regulated environment in order to identify and, if necessary, automatically refine (or prompt refinement of) the system's operations to attain a pre-specified goal

The logic of algorithmic regulation: a taxonomy

- Antecedents in the interdisciplinary science of cybernetics (post WWII)

- Move away from linear understandings of cause-effect relationships towards investigations of control through circular causality or feedback

- Control systems have 3 core components: mechanisms for

a) Standard setting:
b) Information gathering and monitoring
c) Enforcement and sanctioning (to bring behaviour in line with system standard when deviation identified)

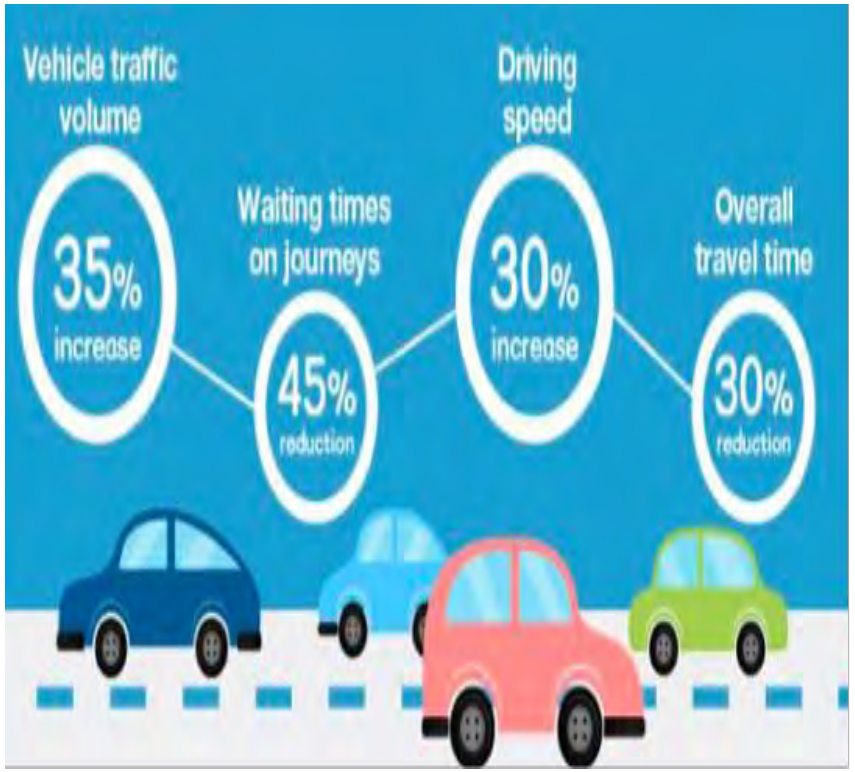
A taxonomy of algorithmic regulation

My taxonomy: two alternative configurations for each component, thereby generating a total of 12 different forms;

Cybernetic component		
Standard setting	fixed ('simple')	variable ('smart')
Monitoring & information gathering	Historic data ('reactive')	Inferred data ('predictive')
Enforcement & sanctions	Automated	Recommendation ('persuasive')

Standard setting

standard setting: behavioural norm either *fixed* ('simple') or *variable* ('smart')

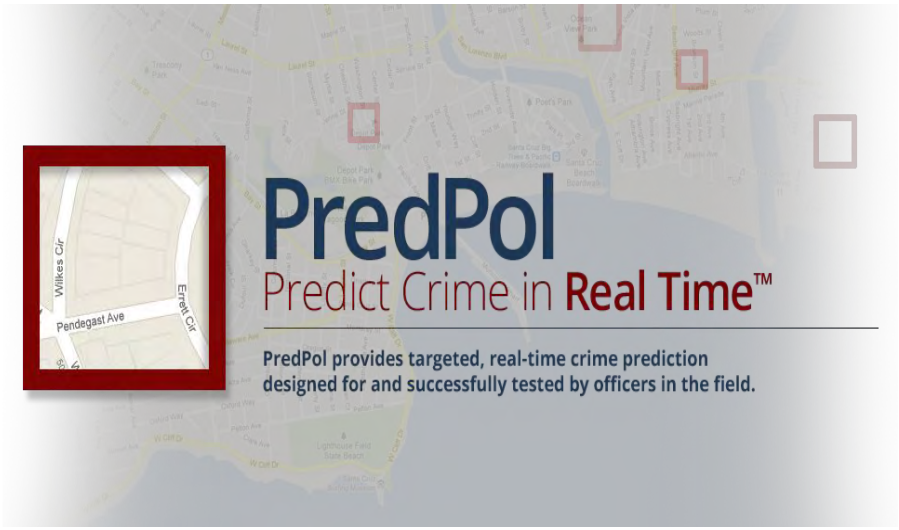


Information gathering and monitoring

information gathering and monitoring: may operate on a **reactive** basis (tracks historic performance data in near real time to detect violation) or detect violations on a **pre-emptive** basis, applying machine learning algorithms to historic data to infer and thereby *predict future* behaviour

Enforcement and sanctioning

Sanctioning and enforcement: administration of decision or sanction may be **automatic** *(*eg password protected access systems) without the need for human intervention, or as **recommender (** 'persuasive') systems, configured to provide automated 'assistance' or 'recommendations' to a human agent, by prioritising candidates from within the larger regulated population, offering prompts that focus a human user's attention on a particular set of entities within the data set, with the human agent retaining formal decision-making authority



Two basic forms: reactive vs predictive

- **Reactive systems**: which trigger an automated response based on algorithmic analysis of historic performance data in real time
- **Pre-emptive systems**: which act pre-emptively based on algorithmic assessment of historic data to infer predictions about future behaviour

These two forms (roughly) track to two important developments in the computerisation of regulatory governance systems which help distinguish what is genuinely 'new' about algorithmic regulation:

- **Automation**
- **Machine learning algorithms fed by large data streams**

Algorithmic regulation as design-based control

- **Automation**: the computational turn makes it possible to automate regulatory governance systems, such as digital password protection systems. Same logic as primitive control systems (eg water lock). Interesting because costs of digital storage and sensor technology so cheap that these automated systems are now *practically feasible*

The novelty of Big Data driven algorithmic regulation

- **Big Data:** machine learning techniques that operate on large multiple sources of data collected from ubiquitous digital sensors that continuously track behaviour, offer a genuinely novel form of design-based control. They enable **'smart'** forms of algorithmic regulation which are configured to optimise a fixed (but reprogrammable) overarching system goal while allowing variation in behavioural standards

- **Population wide reach + concurrent personalisation:** Networked algorithmic systems are vastly more powerful than traditional forms of architectural regulation (cf speed hump or door lock) b/c now possible to track and intervene in the behaviour of a **single user** and an **entire population of users** across a widely dispersed geographic area, while collecting and analysing population-wide data on an almost instantaneous basis to identify deviations from the system's goal

- **Predictive capacity:** machine learning enables prediction of individual and population-wide trends that can reveal, and automatically act upon, 'hidden' insight. It is this capacity to **predict** future action or behaviour based on the algorithmic identification of unexpected correlations within massive data sets that would not be detectable by human cognition (or even ordinary computing techniques) to generate '**actionable insight'** that is widely regarded as the 'Holy Grail' of Big Data.

Reactive algorithmic systems

Reactive algorithmic systems (simple or smart) utilise the logic of traditional **performance/outcome-based management systems** but with three claimed advantages:

- By replacing the need for human oversight with ubiquitous, networked digital sensors, algorithmic systems enable the monitoring of performance against targets at greatly **reduced cost and human effort**.
- They operate dynamically, continuously fed by real-time data, allowing almost immediate intervention to direct or constrain the targeted behaviour, and thereby avoiding problems arising **from out-of-date performance data.**
- They appear to be based on **objective, verifiable evidence** because knowledge of system performance is based on data is collected directly from a multitude of behavioural sensors embedded into the environment, thereby holding out the prospect of 'game proof' design.

The logic of pre-emptive algorithmic systems

Pre-emptive algorithmic systems also offer these advantages, but operate on different underlying logics:

- As a form of **risk-based regulation**: popular with UK governments. Core idea - rather than attempt to prevent all possible harm, regulatory intervention should focus on controlling the greatest possible threats to achieving regulatory objectives, as determined by ex ante assessments of their probability and consequence esp to inform the allocation of enforcement attention and resources – to identify those most 'at risk' of violating regulatory standards

- As form of **actuarial justice**: a theoretical criminological model that employs actuarial mathematics to manage future risks. Aim is not to transform individual criminals but instead to manage risks according to dangerousness of offender, determined via actuarial methods. Assumes that we cannot eliminate crime, so aim is instead to reduce it to tolerable levels, thus reconstructing individuals as risk objects. Primary outlook is prospective, to estimate and prevent the occurrence of future risks rather than sanctioning offenders or addressing past causes

- As a form of **surveillant driven social sorting:** tasks of filtering and classification based on risk assessment long used in the insurance industry to assess individual applicants, but also used more widely by the marketing industry from the early days of data mining in order to target potential customers more effectively, segmenting them into different user groups by profiling individuals.

Algorithmic regulation as a system of social ordering

- Anneesh (2009) identified **'algocracy'** as a system of governance based on 'rule of the algorithm' in his ethnographic study of the labour practice of 'off shoring' through which Indian workers provided IT services to US firms. He identified software programming schedules as critical to the organisation of globally dispersed labour.
- He argued that algocracy is a distinct from both **bureaucratic** and **market-forms** of governance, its underlying logic is driven by the algorithm, which is ***mathematical***

This understanding of algorithms as a **distinct form of social ordering** is an enormously fruitful perspective to ground critical examination:

(1) Politics and drivers
(2) The rise of algorithmic power and the 'new surveillance' – legal and democratic concerns

The Politics of Algorithmic Regulation: Dataism and its Discontents

- Yuval Harari coined the term ‘Dataism’
- an ‘emerging religion’ rooted in a belief that humans can no longer distil the immense flows of data into information, knowledge or wisdom, so that the work of processing data should therefore be entrusted to computational algorithms, whose capacity far exceeds that of the human brain
- **transcends conventional political ideology?** free market capitalism and state controlled communism as merely competing data-processing systems. Capitalism uses distributed processing, by directly connecting all producers and consumers to one another, and allowing them to exchange information freely and make decisions independently vs communism relies on centralised processing.

But this assumes that political systems are only concerned with the optimal distribution of society’s material resources, and crudely overlooks their underlying **politics, values, and normative premises**

The Politics of Algorithmic Regulation

But algorithmic systems have been associated with two dramatically opposed political visions:

- **O'Reilly**: seamless, fully automated data-driven governance that solves societal coordination problems efficiently

- **Morozov**: seeks to expose the hidden anti-democratic vision of Silicon Valley's belief that technological innovation can solve social problems efficiently simply by harnessing the power of the internet

Morozov's critique

Springs from observation that the *means* by which we seek to govern has inescapable political and ideological dimensions, and these shape our substantive political goals. But the politics and ideology of algorithmic regulation are hidden – unlike debates about state vs market, where the ideological dimensions were readily apparent. He identifies '**solutionism**' as the ideology underpinning algorithmic regulation, characterised by the following:

- **Govern effects not causes**: rather than seek to address causes, just seek to manage the effects (movement identified by Italian philosopher Giorgio Agamben)

- **Expand oversight** and collect as much data as you can (not the small libertarian state)

- Encourage individuals to take responsibility for problems (resonates strongly with Foucault's 'governmentality') – individuals responsible for their own health, safety productivity etc via smart tracking devices. Currently portrayed as an optional extra by insurance companies in return for discounts, but in future, failure to track = deviance = higher premiums or even exclusion?

- Characterise **individuals as entrepreneurs** and the **sharing economy as the new welfare state**: individuals are stockholders in a giant enterprise, empowered to take care of their own affairs via ubiquitous digital feedback loops. No assumed social evils that can only be tackled by collective action.

Economic drivers of algorithmic regulation: the emergence of 'surveillance capitalism'

- Combined with the logic of capitalism, this 'Solutionist' mindset is fostering the proliferation of algorithmic systems, driven by a powerful logic of '**Surveillance Capitalism**', a new form of information capitalism (Zuboff 2015)
- Driven by Silicon Valley hyperscale technology companies (and spearheaded by Google) that achieve growth primarily by leveraging automation via global digital platforms
- On this logic, revenues depend upon data assets appropriated through ubiquitous automated operations constituting a **new asset class** (surveillance asset), generating a new default business model where company valuations routinely depend on 'eyeballs' rather than revenue as a predictor of profitability, channelling and controlling flows of personal information while converting them to flows of profit, all in ways that are **highly opaque** to their users.

The New Surveillance

- Surveillance capitalism relies critically on highly granular, population wide and continuously updated **mass surveillance**, all made possible by the networked digital infrastructure upon which industrialised societies increasingly rely

- But unlike the repressive forms of visual surveillance (Orwell's Big Brother), a distinctly western, democratic type of surveillance society emerges. Core logic: we **willingly allow** ourselves to be subject to algorithmic scrutiny by exchanging our personal data for efficiency and convenience

- Rests on a legal model of '**privacy self management'** which legal scholars have trenchantly criticised as inadequate in a big data environment, given the practical impossibility of providing meaningful, voluntary consent to the data sharing activities entailed by algorithmic systems

The New Surveillance and its consequences

- The underlying logic of exchanging privacy for convenience also operates at the collective level. eg NHS Royal Free NHS Trust has agreed to grant Google's Deep Mind access to 1.6 mil patient records to develop healthcare analytics
- Our consent to this exchange may be more akin to that of the compulsive gambler who finds it impossible to kick the habit, despite its harmful long term costs
- For Zuboff, it is 'Faustian pact'
- As Davidow observes, millions of people are now virtually incarcerated in '**algorithmic prisons'** (Davidow 2014) with many people unable, due to their internet profiles, to find employment or who have difficulty accessing various services including insurance, loan finance, rental housing, properties to purchase or to board an airplane

Constitutional values increasingly strained

Legal scholars echo these concerns, highlighting how algorithmic decision-making systems may antagonise constitutional and democratic values, such as

A) **Transparency and accountability**: algorithmic processes are highly opaque and impossible for the lay user to comprehend. Individuals may be unaware that these automated processes are in use, or even what kind of behaviour or trait condemned them in the first place

B) **Informational privacy and fundamental rights**: threats to the right to informational privacy most obviously threatened by algorithmic regulation and individual autonomy and self determination more generally. But not easy to fit these concerns within **legal rights discourse** (due to the way in which algorithmic 'profiling' operates – not necessarily correlatable to a biographical individual)

C) **Due process and rights of appeal and redress:** typically no mechanism through which an individual can contest algorithmic decisions, hence a threat to due process (that those affected by governmental decisions should have an opportunity to participate in them). Coupled with rising anxiety about the freedom of powerful firms to act unilaterally against individuals without giving them an opportunity to contest or challenge such action

a) **Equality of treatment:** much anxiety about the capacity for algorithmic systems to discriminate against historically marginalised social groups . As Oscar Gandy warned back in 1992 that data mining technologies are 'discriminatory by design', and hence risks undermining key aspects of democracy, equality, fairness and distributive justice

Algorithmic power

- **Power asymmetries:** Eg WEF (2014) recognises that governance issues need to be addressed in order to protect the rights and claims of individuals, particularly given that the lack of power of individuals is a serious challenge, in which power currently favours institutions (both public and private) with individuals largely passive data subjects who lack meaningful influence and control over the disposition and use of their personal data

- Hence the WEF argues that we need to ensure that the algorithms that drive these anticipatory decisions will **be 'lawful, fair and can be explained intelligibly'** – but what does this require?

Automation and the distribution of decision-making authority

Might keeping humans in the loop (by configuring as 'recommender' systems rather than automating enforcement) overcome these concerns? Seems highly **unlikely**

- Merely keeping a human in the decision-making loop does not itself satisfy the demands of due process
- Even if a human retains formal authority to make a decision, that authority can itself be improperly exercised or abused
- Affected individuals may be unaware that they have been subject to an adverse decision based on algorithmic evaluations that may have no causal basis whatsoever: but this lack of awareness does not legitimise the decision
- Humans are highly susceptible to 'automation bias' – ie tend to defer to computational judgments even when capable of recognising that the situation calls for another choice
- Algorithmic recommender systems can be a very powerful forms of choice architecture that can manipulate in subtle but highly effective ways ('hypernudge')

Algorithmic accountability and the social foundations of democracy

- Recognition of the rise of algorithmic power and its inscrutable processes (algorithm as 'black boxes') is driving demands for '**algorithmic accountability**', highlighting the need for mechanisms through which algorithmic decisions can be explained and justified to those affected against some criteria, and to make amends for any fault or error
- The need for **explainability** is esp acute in liberal democratic societies – ie those which aspire to be transparent orders, in that its workings and principles should be well known and available for public apprehension and scrutiny, so that the social order can be justified to those who live under it (Waldron 1987)
- Concerns about the risks to collective values of **transparency and accountability** highlight how a wholesale shift towards algorithmic decision-making systems risks eroding the **collective moral and cultural fabric** upon which democracy and individual freedom rests
- Privacy is not merely an individual right, but a **collective good** – it provides a zone of protection around each individual's activities within a society, making possible the capacity for individual flourishing and self-creation, in which our sense of self and our individuality can emerge, mutate and stabilise (Cohen 2012). Without it, there is no democratic or individual freedom: yet the importance of this critical moral and social infrastructure is frequently overlooked in contemporary debates

Constitutional democracy as a cybernetic system

Contrast the legitimacy and logic of contemporary algorithmic regulation (in the form that is currently emerging) with constitutional democratic systems of governing:

:

- Mireille Hildebrandt: constitutional democracies render sovereign rule legitimate via a double form of transparency:
- People live under rules of their own making (democratic participation
- The application of those rules can be contested via a procedure that is capable of opening the black box of their interpretation (the rule of law

- This generates a cybernetic system, ie constitutional democracy, in which all who live under the rule of law are not objects to be controlled, but subjects participating in collective self rule and accountable to each other, and to their government.

Conclusion

- Algorithmic regulation, in the form that it is currently emerging in contemporary modern democracies, threatens this constitutional balance – it provides a **one way mirror** that allows institutions looking down to surveil those below, yet those below lack any prospect of peering, let alone understanding and challenging these algorithmic black boxes that regulate their lives

- But, **it need not be so:** these choices are **political** (albeit powerfully shaped by economic drivers). It is theoretically possible to envisage more egalitarian, progressive systems of algorithmic regulation.

- For example - algorithmic systems could operate as **two-way mirrors**.
- customers could monitor the performance of corporate service and government providers, as their own adherence to laws and contract terms are monitored.
- individuals could simultaneously monitor the actions and performance of their employer (not necessarily confined to human resources issue handling) and governing institutions whilst their own productivity and performance is individually monitored.

Conclusion

Yet, these are not the kind of algorithmic platforms that Silicon Valley start-ups are keen to develop: for they do not offer the lucrative financial returns that accrue to giant digital platforms that wield asymmetrical power vis-à-vis their users. So, I doubt that systems of this more progressive, egalitarian kind will emerge spontaneously from the capitalist market order without political intervention.

Hence a core challenge lies in the political realm: how can we foster active, meaningful debate and deliberation that can shape, inform and constrain the way in which algorithmic systems are developed and implemented, that will reflect the core values and aspirations of the populations which these systems increasingly regulate?

Contact details
Karen.yeung@kcl.ac.uk

Thank you.