Artificial Intelligence, Big Data and the Rule of Law

Event Report

Date: 9 October 2017

Venue: The Law Society

Speakers:

- **Iain Bourne**, Information Commissioner's Office
- **Silkie Carlo**, Liberty
- **Prof Lorna McGregor**, University of Essex
- **Marion Oswald**, University of Winchester
- **Prof Ian Walden**, Queen Mary University of London

Chair:

- **Christina Blacklaws**, Vice President of The Law Society

Organisers:

- This event was organised by the Bingham Centre for the Rule of Law in partnership with The Law Society. It was convened by Lucy Moxham (Associate Senior Research Fellow, Bingham Centre for the Rule of Law).

The programme is available here.

Christina Blacklaws made some introductory remarks, stressing that no contemporary discussion of the rule of law could ignore the growing impact of technology in transforming the legal landscape. She noted that in his 2010 book 'The Rule of Law' Tom Bingham referred to a study by Privacy International which found that the United Kingdom was the most closely watched country in Europe and noted that this was made possible by the notable technological advances in recent years. These changes have accelerated since then and Ms Blacklaws emphasised the necessity to balance the potential social benefits against possible dangers such as algorithmic bias and discrimination. The rule of law ought not to be displaced by the rule of computer code. Ms Blacklaws emphasised that we need to understand and debate the legal, ethical and moral implications of these powerful technologies that are developing so rapidly.

In his opening remarks, Iain Bourne gave an overview of some of the challenges faced by regulators with the spread of automated decision-making. Mr Bourne noted that one of the policy drivers for next generation data protection laws (e.g., the forthcoming General Data Protection Regulation (GDPR), which will apply in the UK from May 2018) was concerns about automated decision-making processes. He commented that this is becoming a more live issue for law-makers, regulators, lawyers and technologists and highlighted the increasing awareness among the general public about the impact of such systems. As regards to the role of the Information Commissioner's Office (ICO) in this context, Mr Bourne noted the increasing use of automated decision-making in areas which is already having a significant impact on individuals' lives e.g., in the employment sphere, counter-terrorism checks, parole decisions etc. Even though data protection law is about to be modernised and upgraded, existing legal protections do not fit with the rapidly evolving technology.

First, the traditional data protection model used by regulators which is based on principles of fairness and reasonableness is under strain in light of the new technology and we need to develop new ways of regulating this technology in the most effective way. For example, there is a real issue of data controllership with AI-enabled kit which can essentially decide itself what data sources it is going to use to collect information about people and how it is going to analyse it. In this case, it is uncertain what the outcome is going to be, and if and how unfair decisions, bias and discrimination can be programmed out from AI-enabled systems. Second, in this regard, Mr Bourne asked how these systems can be controlled and he emphasised that this is an issue for technologists as well as lawyers and regulators. He highlighted the lack of understanding amongst regulators and lawyers as to how these systems work, and he suggested the need to boost technology teams to provide such knowledge and to improve resilience. Third, he asked how we can programme fairness into these systems. He noted examples of AI-enabled systems in a "race to the bottom". While AI and automated decision-making are capable of guarding against human-generated discrimination and bias, they are also capable of creating their own.

He explained that the new data protection rules will have a lot more emphasis on automated processing including the use of AI, rules about transparency, rights to stop the processing of personal data by automated means and protection via an element of human supervision etc. While we already see human supervision e.g., in credit reference decisions, he pointed out that the possibility of providing any kind of effective human oversight is under strain given the sheer scale of data involved in e.g., social media content moderation and online dispute resolution in e-commerce. What can citizens expect big social media platforms and technology companies to do in these situations? How do regulators explain how AI works to the general public?

Finally, Mr Bourne spoke about creating a more ethical approach to data usage. He noted that consideration is being given to setting up data ethics bodies and suggested that encouraging a more ethical approach to data usage fits well with this AI set of issues. He noted these discussions might be based around fairness, appropriateness, social value and societal impact etc. There is a lot of work being done on data ethics at the moment. So we may need to think about data usage in a wider sense than just personal data and look at other aspects of human-computer interaction. A set of value-based principles around data ethics to deal with the "should we do it" question would be interesting here. Mr Bourne emphasised that while there could be a number of bodies forming a diverse community for ethical oversight, but the existing roles of regulators and others must be respected. This is a fast-moving space and the government is clearly very interested in data ethics, as this would help them to give the public reassurance about data usage and safeguards.

For further information, see the ICO's recent report on 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' here.

Silkie Carlo explained that she would focus on how big data fits into or challenges the human rights framework, in particular privacy, discrimination and liberty. The growth of data and the increasing ability to store and process it are reshaping society. Big data presents a real risk to citizens' privacy. Personal data is the primary concern from a human rights perspective and it is now a commodity which is driving the technological revolution. As such, lawyers, technologists and civil society have a privilege as well as a challenge of being at a unique and vital axis in time when precedents set today may outlive us and help to uphold the protection of rights and the rule of law in the course of the technological revolution and beyond.

Concerning the right to privacy, Ms Carlo noted that big data is increasingly seen as central to national security, law enforcement and even the provision of services and access to those services. Data visibility, particularly in relation to the state and in the name of national security is a growing norm that will renegotiate the relationship between the citizen and the state, and the way we experience our fundamental right to privacy. She discussed the December 2016 judgment of the Court of Justice of the EU in Tele2 Sverige and Watson and Others (conjoined cases C-203/15 and C-698/15). She noted that since then and despite the clear legal issues and the concerns raised by civil society, the Investigatory Powers Act has now been enacted.

Ms Carlo observed that it seems increasingly that the notion of a private conversation beyond the eyes of the government is per se dangerous and subversive, and her concern that this seems to be government's starting point for civil liberties in the digital age. She also mentioned as a further issue the increasing use of biometrics which is also posing new human rights challenges, in particular to privacy. This represents a new invasion into the private, personal sphere and this intrusion is part of proving one's continued innocence in public spaces. She gave an example of the facial recognition software trialled by the Metropolitan Police at the Notting Hill Carnival in 2017 (see e.g., media coverage here). Ms Carlo emphasised how the loss of privacy is reshaping social spaces and that it is a right that is essential for democratic processes to prevail. If law and policy are not catching up, it may be necessary for civil society to bring innovative legal challenges.

Ms Carlo then moved on to the risk of perpetuating discrimination. Big data represents complex social histories, discrete biases and patterns of discrimination. If we feed AI tools and algorithms with data unconsciously and tell ourselves that data is neutral and objective then we will perpetuate discrimination and inequalities that we are consciously trying to eliminate in human decision-making. Given the volume of data and complex analytics involved, she noted that there are many more vectors by which discrimination can occur and with more subtlety and less direct accountability. She noted that such discriminatory biases can interfere with a broad range of rights and freedoms.

Ms Carlo noted that there is already evidence that data science has perpetuated discrimination in the criminal justice system in the United States and she highlighted the example of an algorithm called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) in that regard (see e.g., media coverage here). She also noted that Durham police is using similar artificial intelligence software in bail decisions as well (which is discussed by Marion Oswald later in the programme).

In her view, the threshold for shifting to data-driven and especially algorithmic-driven criminal justice and law enforcement should not simply be whether one function of a software exceeds human functioning on some measure, but moving to the software should be considered holistically accounting for new risks as well and long-term impact. She stressed that data analysts should at very least attempt to discover and control the biases in existing data sets before using them to train AI tools or live deployment in the criminal justice system where they risk being embedded and obscured from accountability. Seemingly progressive AI applications may actually reduce socially progressive outputs so we must not automatically attribute them as objective or divorced from ideology.

Lastly, Ms Carlo spoke about liberty and in particular how citizens can become suspects in the big data context. She noted that analysts have suggested that AI and big data technologies are going to transform policing in the UK, potentially both in terms of the tools they use and principles upon which policing is based. She explained how already almost all activities in our modern lives can leave a data trail that can be followed. For example, she noted the use of information from a smart water meter and from an Amazon Echo device in a murder case in the US (see e.g., media coverage here).

Ms Carlo noted her concern with the increasing availability of big data and its use in forming suspicion, that as it sinks into public consciousness that the police are increasingly using this big data, a chilling effect will be felt across society. She noted that experts have argued that it is restructuring police–community relations whereby citizens are losing a measure of liberty. She suggested it is important for us to consider what the big data context is doing to the very notion of reasonable suspicion. She suggested the threshold of reasonable suspicion is being subverted or may become practically irrelevant in an era of big data policing. She noted that the threshold for reasonable suspicion is traditionally met by an individual's observable behaviours, however in the modern context reasonable suspicion derives from an individual's obtainable data. She commented that this represents a radical departure from the original premise of reasonable suspicion. Finally, Ms Carlo noted that the impact that big data then could have on stop and searches which are already renowned for their propensity to discriminatory targeting, could be extraordinary. She noted the potential that non-criminal actions could satisfy the reasonable suspicion standard. She suggested that replacing the need for observable suspicious behaviours with obtainable suspicious data points could lead to arbitrary and speculative stops that target innocent people.

For more information, see Ms Carlo's presentation here.

Prof Lorna McGregor elaborated on the opportunities and challenges posed by artificial intelligence technology and big data for human rights. She underlined that there are issues here for all our human rights, not only for privacy. Privacy plays a critical role and is the linchpin to the enjoyment of all our rights and even what is means to be human. If privacy is at risk of being violated, it can have a chilling effect on the way we live. She emphasised that it is not about pushing back on the use of big data and artificial intelligence, but rather about trying to understand the human rights implications and opportunities. To illustrate this, she mentioned the UN Sustainable Development Goals (SDGs) and it is often said that big data and artificial intelligence are crucial to the advancement of the SDGs and that we need to ensure that no one is left behind. So another part of this debate is about ensuring that big data and AI are not used solely to the benefit the global north or certain companies, but that we all share in the benefits. Prof McGregor emphasised that the goal is therefore about understanding the risks and ensuring that a human rights framework is in place to protect against these risks while ensuring that we can benefit from innovation that is taking place and that is going to continue to take place.

Prof McGregor then focussed on the impact of algorithmic decision-making and algorithmic accountability. She noted that there are huge human rights implications brought about by algorithmic decision-making. With algorithms themselves becoming increasingly complex and sophisticated, and with advances in technology, we are moving beyond algorithms that are easy to understand. Coupled with automation and the possibility of autonomous decision-making, again changing the nature of the algorithmic landscape and with that interaction with big data.

She then discussed predictive policing and noted that while algorithmic decision-making is on the one hand a useful tool for police when they are considering how to allocate resources to help with crime reduction, other studies show concerns that this will then exacerbate existing inequalities in policing and existing discrimination where certain offenders and certain communities are over-policed or discriminated against. Prof McGregor also noted the US case where algorithms were being used in judicial decision-making. Here she underlined not just the way in which these studies about risk assessments are undertaken and the risk for discrimination within them, but also that there was an attempt to challenge the use of algorithms for risk assessment, and the response there that the outcome was sufficient, based on the idea that the technology was more predictable and neutral in some way. So we also have to understand that technology can be as fallible as humans. Prof McGregor also noted a recently reported study suggesting that facial recognition software can be used to identify an individual's sexuality (see e.g., media coverage here). Putting to one side issues around accuracy, she noted that this also raises questions about whether it is appropriate to be using algorithmic decision-making in all circumstances or whether there are red lines we ought to be considering in terms of where we should not be employing this kind of software.

Prof McGregor noted that these examples reveal a wide range of human rights issues, from the impact algorithms have on liberty, to the effect on surveillance, privacy and the chilling effect. Discrimination is a huge risk here both through feeding algorithms with discriminatory data and also with algorithms themselves acting in a discriminatory way. She also noted inequality and discrimination in terms of who is subject to algorithmic decision-making and who still gets access to a human decision-maker, and asked where the checks and balances, and fairness are in these processes. She raised further questions as to the ease of challenging algorithmic decision-making and here she highlighted the US case noted above where whilst there was the possibility of challenging the decision, but there seemed to be a strong trust in the objectivity of technology and of algorithmic decision-making. Even though there was an element of human involvement, we also need to question how meaningful that is when there is algorithmic decision-making taking place. This is particularly the case where algorithms are becoming more autonomous, complex and difficult to understand.

Lastly, Prof McGregor outlined some ideas as to how we might move forward. Whilst we are late in coming to the table, it is not too late to think about what regulatory frameworks might look like and for example what multi-stakeholder frameworks mean at the national, regional and international level, in order to address some of the human rights impacts we are seeing. She also underlined that while the technology is out there, some of these current negative effects can be rolled back. The starting point is to understand the bigger situation that we are facing, which is a world which is driven by big data, algorithmic decision-making and the increasing presence of artificial intelligence, and how they are interdependent and intersect.

Prof McGregor noted that there are many discussions taking place about ethics and ethical approaches, about how to make algorithms more transparent, how to deal with proprietary interests in algorithms etc. There is a lot of debate about whether the GDPR is requiring explainability of algorithmic decision-making, and questions around trust and fairness. All of these elements are important when we are thinking about algorithmic accountability. However, Prof McGregor emphasised that when we focus on ethics, we also need to remember responsibilities. We need to consider what the business and human rights framework means in this context. Are we looking for voluntary engagement or for a different model? We also need to think about how responsibility works across the algorithmic life cycle, especially when there is self-learning involved. In other words, when is the developer responsible and how does it work throughout the life cycle of an algorithm.

Prof McGregor emphasised that while the ethical approaches are key to dealing with these issues and an understanding of the responsibilities of businesses and states are crucial, we also need to look at the existing human rights frameworks in this context. She noted that with international human rights law we already have a framework which considers prevention, monitoring and oversight, remedies and accountability. At the prevention stage, we are hearing discussions about how developers can include ethics from the outset. However, we need to think about what the criteria would be here – what ought to be built into an algorithm and are there any red lines where algorithmic decision-making ought not form part of a particular public or private decision given the risks to human rights. There are also questions around the design of impact assessments – how can we design impact assessments that are ongoing, that can trace an algorithm that is changing to see whether there are any unintended human rights consequences, how can that be monitored, and how do we design oversight bodies that can work with proprietary interests.

Lastly, Prof McGregor highlighted the question of remedies – from a human rights perspective we also need to ensure there are models of remedies that work for individuals and groups.

Marion Oswald concentrated on algorithmic policing, with a particular focus on the tool currently being used by Durham Constabulary, the Harm Assessment Risk Tool (HART). Ms Oswald began by emphasising that in reality many police forces currently lack the technological capability to use digital data effectively. However, she noted that this is changing and developments are happening in three main areas in a policing context: predictive policing on a macro level; operational, e.g., intelligence-linking in conjunction with an investigation; and decision-making or risk assessments relating to individuals. The HART tool is an example of the third type of algorithmic policing.

Ms Oswald emphasised the importance of considering the context of each algorithm. For example, the context of the HART tool is Durham Constabulary's 'Checkpoint' programme which is an out-of-court disposal aimed at reducing re-offending and improving the outcomes for victims. She explained that the programme considers offenders arrested for relatively low-level offences where there is sufficient evidence to charge and instead of going through the normal court system, routes them through the Checkpoint scheme, which aims to tackle their individual issues (such as alcohol and drug problems etc.) and by giving them the opportunity to sign up to certain actions to reduce their chance of re-offending. She noted that these types of schemes were recently mentioned in the Lammy Review, which praised the Checkpoint scheme in Durham and recommended that there should be more of these types of deferred prosecution schemes in order to ensure more overall justice in the criminal justice system.

Ms Oswald then gave a more detailed description of the Checkpoint scheme. She explained that there needs to be a decision made as to the eligibility of a particular offender for the Checkpoint scheme. There is a risk forecasting algorithmic model which is used to support decision-making by the custody officers who are the ultimate decision-makers. They are thinking about whether someone is at high risk, medium risk or low risk of re-offending. Checkpoint is targeted at those offenders considered to be at medium risk. The tool uses a random forest machine learning approach, which is a way of using many independent decision trees to get a combined and potentially better result. The major issue around these tools is that past behaviour is no guarantee of future behaviour. In this specific tool, Durham uses 34 predictors (about which it has been transparent), which include age, gender, postcode, age of first offence, type of offence and criminal history. Ms Oswald noted that Durham is particularly concerned to deal with false positives (predicted as high risk, but turns out to be a low risk offender) and false negatives (predicted as low risk, but turns out to be a high risk offender). From a policing perspective, false negatives are particularly concerning errors.

Ms Oswald noted that the HART model raises all the issues mentioned by some of the other panellists and others – For example, the output is probable but not conclusive so how do we ensure that is dealt with appropriately; judgemental atrophy whereby police officers take the decision of the algorithm and do not apply other judgment to it; opacity and how these algorithms explain their decisions so they can be interrogated; risk of bias; certain data sets such as residential address could be proxy for protected attributes; and ultimately we need to consider whether these tools are necessary, proportionate and in accordance with the law, both as to their means and their ends. However Ms Oswald also noted that these sorts of tools do have potential benefits – focusing on consistency in decision-making by combining the experience of many custody officers; enabling testing and adjustment to be done in a systematic way; and if done better it could make the decision-making process more transparent. In this specific context, effective forecasting can lead to more effective outcomes for offenders. She underlined that risk assessment in the policing context is hugely difficult. In her view, the issue of false positives, false negatives and trade-offs raises one of the most difficult issues in this context (which is the more dangerous or problematic error). The other big issue is the "computer says no" risk.

Ms Oswald concluded by emphasising that such tools in all sorts of policing contexts are effectively experimental and the long-term benefits, harms or outcomes are not yet clear. She stressed that the public sector needs to be able to think about the ways of doing things better in relation to these new technologies and the public needs reassurance regarding oversight and regulation. For this reason, in the context of her work with Durham Constabulary, they have proposed two linked models: a model of experimental proportionality which permits the use of unproven algorithms by the public sector in a formally regulated and time limited way, in combination with a robust decision-making model.

For more information, see Ms Oswald's presentation here.

Prof Ian Walden began by emphasising that we need to consider the system as a whole. Algorithms, big data, connectivity, and the workflow process within which machine learning or artificial intelligence applications are deployed – He stressed that all these components are all opportunities for us to exercise control, to impose liability, and to use traditional legal techniques to regulate artificial intelligence and big data. Part of this of course is about decision-making and Prof Walden highlighted three key points in this respect.

First, he spoke about the representative nature of the data and emphasised the importance of training data. He explained that whether it is supervised or unsupervised learning, machine learning and AI applications have to learn on data and therefore the data that is supplied to them, whether in terms of quality or quantity, becomes a critical differentiator in terms of how the law treats a particular AI application.

Second, in terms of discrimination, Prof Walden noted that while discrimination can be built into such systems, they also have the ability to identify bias and variance. He commented that the law has a peculiar way of addressing questions of discrimination – there are examples of correct decisions which are incorrect because the law decides that they are inappropriate (in this respect he noted the example of the inability of insurance companies to discriminate on grounds of sex – see e.g., discussion here).

Third, concerning the question of transparency, Prof Walden noted that we have to decide transparency to whom. There is transparency to the operator themselves and whether when they deploy an AI application they have the required transparency to exercise control and if not, what might the legal system do in terms of imposing liability. In terms of third party transparency, it is not just about the users, it could just be the regulators in certain circumstances. There is the possibility that transparency does not need to be shown to the end user as long as there is regulatory oversight that can impose liabilities and remedies in the event of wrong. As with any technological development there needs to be a balance of interest and that balance of interest is evident in the GDPR to the extent that it says that trade secrets may need to be maintained and we may need to impose an accountability obligation on those who deploy AI applications in terms of being able to extract the rule that gives rise to the decision. Finally, on transparency, he noted that we need to ask whether we want to impose ex-ante obligations about the design of these systems (designing into the system a rule extraction mechanism) or ex-post transparency obligations (with mechanisms for understanding why a particular decision was made).

Prof Walden then moved on to the question of liability – What is the basis for allocating responsibility? Do we want to perhaps impose strict liability? He noted that there has been a call for strict liability in respect of software as a general topic, though the government has resisted this. But he suggested that perhaps we are starting to see in particular software applications, with AI being a subset of that, where government will accept some sort of strict liability. He pointed, for example, to section 2 of the recent Vehicle Technology and Aviation Bill which imposes strict liability. An alternative form of liability is negligence; however, Prof Walden observed that this gives rise to a number of problems such as causation. He suggested that in the absence of an accountability mechanism built into the AI application and if you are not able to account for the way the decision has come about, the methodology, the data and the process, then the evidential presumption is that your application has caused the harm and you will be responsible. Prof Walden noted that another problem under negligence liability is the standard of care problem – What should that standard of care be? Should it be reasonable? Should it be appropriate? And how can we look to standards and certification systems as a mechanism of embodying in law responsibility for AI applications that do not meet certain minimum criteria. Prof Walden also discussed attribution of liability in this context. Next, Prof Walden considered potential remedial mechanisms – Do we want to perhaps for example shift the remedy to the insurance industry? Alternative mechanisms could exist where we essentially remove liability. He stressed in this regard that we need to consider the role of regulation and the extent to which there is a problem of law with artificial intelligence.

Prof Walden then explored some questions with regard to the regulation of data. To what extent is more data always better? He highlighted that there are rules under data protection law to try to minimise the amount of data that is generated about us both by design and by default; to try to stem the tide of personal data being exploited in various environments. We also need to think about to what extent we can control the information flows better – he noted there are computer science initiatives about information flow control that would allow us to control better the way in which information flows within systems, whether it be AI or otherwise. In terms of the rule of law, Prof Walden noted that we need to consider questions that AI is not uniquely raising but which are continuing to be problems in a highly technological environment – He noted, for example, issues such as consent, transparency, and whether we have the opportunity to know how a particular AI application is going to operate, which he suggested are to a certain degree fictions that are becoming more and more strained in a technological environment, of which AI is really just the latest manifestation. Prof Walden concluded that while artificial intelligence generates a lot of challenges for the rule of law, the law will not be found wanting. It is simply a matter of applying familiar conceptions either directly or through regulation in this new environment.

This report was prepared by Lucy Moxham, Associate Senior Research Fellow at the Bingham Centre for the Rule of Law, with assistance from Anja Bossow a Research Volunteer at the Bingham Centre for the Rule of Law.