Catching up with the Debate: Artificial Intelligence & the Rule of Law

Ellis Paterson

This piece was originally co-authored with Gemma McNeil-Walsh for the RECONNECT blog .

We are living in an age of profound technological advancements. For artificial intelligence - the development of systems that simulate human intelligence - it is still early days, but development is gathering pace. The opportunities and capacity for artificial intelligence (AI) to transform our public services, the economy, and areas such as medicine and education, are growing exponentially.

The emergence of a world characterised by the prolific use of AI raises a series of important questions which can be considered to fall into one of two categories. First, what are the legal boundaries to AI systems? Are existing legal frameworks fit for purpose in the context of AI, and if not, what should they look like? Secondly, even in the context of what may or may not be legal, what are the ethical boundaries to AI systems? What kind of a society do we actually want to move towards and what do we want for the future of the human condition? These two categories of debate are by no means mutually exclusive, but they certainly raise different spheres of challenge. Taking the context of surveillance, for example, we see that the question 'Is this level of surveillance legal?' is quite different to the question 'Do we as a society of human beings want to be subjected to this level of surveillance?' Throughout this post we will make reference to the European Commission's High-Level Expert Group ('HLEG') on AI's 'Ethics Guidelines for Trustworthy Artificial Intelligence', which serves as a helpful and comprehensive prism through which to channel key AI and Rule of Law (RoL) questions.

AI, ethics and the RoL

In starting to grapple with AI, the ethics debate has arguably been more fully developed and engaged with than the legal debate. In May 2019, the OECD member countries adopted the 'OECD Principles on Artificial Intelligence', outlining five complementary values-based principles for the responsible stewardship of trustworthy AI. These principles were drawn upon by the G20 in adopting the G20 set of human-centred AI Principles. HLEG's Guidelines defines Trustworthy AI as AI which is lawful, ethical and robust.

The HLEG's Guidelines are addressed to all AI stakeholders designing, implementing, deploying and using AI across the private and public sectors. The HLEG believe that AI is not an 'end in itself' but a way to enhance individual and societal progress and innovation. It is through adherence to the HLEG's Trustworthy AI principles that the Group envision European citizens reaping the benefits of AI in a way that is concomitant with respect for human rights, democracy, and the RoL. Most importantly for the purposes of this blog, the HLEG is keenly aware of the ethical challenges that AI raise, such as in decision-making capabilities, safety, and its impact on EU citizens in myriad ways.

The RoL lexicon

The RoL is a principle of governance that introduces to the debate on AI a unique set of concepts such as separation of powers, participation in decision-making, avoidance of arbitrariness and fairness in the application of the law. In February 2019, a high-level conference was held in Finland, co-organised by the Finnish Presidency of the Council of Europe Committee of Ministers and the Council of Europe ('CoE') on the impacts of AI development on human rights, democracy, and the RoL. The RoL is one of the three main pillars that constitute the CoE core values.

The principle of the RoL is about more than making clear, just and accessible laws (although that is, of course, a crucial element). Adherence to the principle means holding the government and private actors to account. It means that citizens should have access to fair and impartial dispute resolution, and should be able to expect that the processes by which laws are made, administered and enforced are fair and accessible. These concepts are particularly useful in framing our understanding and thinking about the governance of AI. The conclusions from the CoE Finland conference on AI place great importance on the fact that a RoL-driven approach to AI development and use is one that requires 'timely and thoughtful policy responses' and interdisciplinary and independent research into 'its direct and indirect effects on individuals and societies in concrete contexts.' We agree with this analysis, given that AI has the potential to reach deep into society across a number of different sectors.

In this blog, we dig into a few of these concepts in more detail: accountability, transparency, and the protection of fundamental rights. In the following examples, it is also useful to recall one of HLEG's foundations for Trustworthy AI: AI systems 'must not undermine democratic processes, *human deliberation*, or democratic voting systems' (emphasis added).

Accountability

One of the most important elements of the RoL is that all persons, including the Government and private actors, are accountable under the law. If AI systems start to make decisions that have been traditionally made by officials (for example, in welfare or immigration), there is the risk of an 'accountability deficit'. Who exactly do we hold accountable if a decision is made by an algorithm? Would it be necessary for it to be traced back to the nearest human decision-maker (such as a Minister), or would accountability rest with the company that produced the algorithm?

A second concern with ascertaining accountability is that if a decision is taken by an AI system, and the processes are completely opaque (meaning that no one, except the programmers, can explain how the machine came to this conclusion), how can citizens ever be in possession of relevant and requisite information to allow for a meaningful and robust challenge of such a decision? This runs the risk of undermining two crucial RoL principles: one, that the law should provide access to justice (especially where people cannot resolve interpersonal disputes themselves); and two, that courts and tribunal processes should be fair. The ability to scrutinize decision-making processes (which is imperative to the ability to scrutinize the outcomes of those processes) is a keyl feature of those justice systems underpinned by fundamental RoL principles. The HLEG Guidelines state that accountability is closely linked to the principle of fairness. While we would typically associate this with fairness of procedure for citizens, the Guidelines note that this principle encompasses the auditability of AI systems. The evaluation of such systems by internal and external auditors is a crucial part of the accountability issue, as it goes some way to ascertaining who and what is responsible for these systems. Furthermore, HLEG suggests accountability via governance frameworks, which may potentially evolve into EU commitments that Member States should adhere to as the use of AI continues to proliferate.

So, in an age of decisions being made by AI systems (whether in full, or in part), what does that scrutiny look like?

Enacting legal regulation through the lens of the RoL may help programmers and data scientists design systems that allow for better ascertainment of accountability (like, for example, mechanisms by which the system can be scrutinized) which may in turn

add to increased transparency in the design and deployment of algorithms.

Transparency

As highlighted above, understanding the processes by which decisions are made is fundamental. To challenge a decision made by a public decision-maker, it is essential to understand how and why the decision-maker reached a particular conclusion. With data and processing, we still need be able to see and understand how decisions are made. If an individual is unable to prove X, Y, or Z (because they either don't understand a particular decision-making process or they do not have access to the data behind the decision) it can be very difficult to make a meaningful complaint. HLEG calls these the elements of 'traceability' and 'explainability,' The former refers to the data used to build the algorithm, and the latter is a requirement that decisions made by AI systems can be understood and traced by human beings. These are important ideas to keep in mind, particularly when considering how AI will change the way we evaluate decision-making.

There are significant barriers to ascertaining how automated decision-making systems are created. For example, the testing and training data that a company feeds into its algorithm is likely to be protected as a trade secret. Such training data is an essential piece of the puzzle in figuring out how a system reaches a decision. Yet without it, citizens - and public bodies - are left in the dark about clarity, intelligibility, and predictability of these systems.

Furthermore, there may be institutional knowledge difficulties: i.e. that public bodies themselves are not clear on how these systems are used, or how they reach decisions. This is symptomatic of officials themselves not understanding exactly how the systems work, as they can be very technical. In the vast majority of cases these systems are not in-house, but are procured. Therefore the data points may be in-house also, which may differ from the data points used by public and administrative bodies. The fact that these systems are incredibly complex and that the vast majority of people don't understand them can be seen to give them an enhanced (but highly problematic) sense of legitimacy. Just because they are complicated and technical, does not mean that they are necessarily 'better' processes by which to make decisions not that they necessarily produce 'better' outcomes.

Adequate protection of fundamental human rights

It is important to remember that what is considered ethical and fundamental is subjective, and often, in the case of public decision-making concerning decisions like welfare allocation, highly context-dependent. A challenge here is how algorithms would incorporate such subjective and context-dependent information into the system, and take into account those changes over time when the system was making a decision.

In the UK, a local police force have started to use an online Harm Assessment Risk Tool in custody suites to process individuals post-arrest. It decides whether an individual represents a high, medium, or low risk, the determination of which will influence the decision as to what happens to that individual (a high risk would see individuals remaining in custody, whereas a low risk will have them released on bail). In an assessment of the Tool, it was found that the decisions of the system, and the decisions of human police officers had an agreement rate of only 56%. Where an individual's assessment produced a 'borderline' result, the system would automatically designate them into the higher risk category. However, human police officers, understanding the implications of an individual being designated high risk, are more likely to choose moderate or low risk where the individual presents a borderline result. A human police officer is only likely to choose high risk 1% of the time if a borderline result is produced.

Under the RoL, the law must adequately protect fundamental human rights. It may therefore be the case that in situations where an individual's fundamental rights are on the line, AI-led decision-making may not be reliable in ensuring adequate protection of those rights. And, if we do not yet have the assurances that AI decision-making or decision-assisting systems do adequately

protect fundamental human rights, should they even be in operation at this stage?

Within the EU, the Charter of Fundamental Rights and EU Treaties setting out fundamental and human rights are legally binding. As one of the bedrocks of Trustworthy AI, the HLEG note that adherence to these rights are legally obligatory by entities involved in the creation and distribution of AI systems. There are ways for AI companies to begin to adhere to such rights. For example, respect for human dignity in an AI-context requires that individuals are treated respectfully as moral subjects and not merely objects to be run through an algorithm. This is the kind of human-centric thinking in formulating AI systems that the HLEG supports. While a legal challenge could be raised after the fact, the issue remains however, of ensuring that these standards are implemented at the point of creation.

Further, the CoE Commissioner for Human Rights, in May 2019, published a report on protecting human rights within the domain of AI . This raft of recommendations includes the promotion of AI literacy amongst citizens of Member States, obligations of Member States to facilitate the implementation of human rights standards in the public sector, and advocates for the prevention and mitigation of discrimination risks. We believe this is a valuable set of recommendations that not only maps out issues concerning AI and fundamental rights, but provides viable solutions for the Governments of Member States.

Conclusion

In this blog we have illustrated just a few examples of the concerns that AI and automation present for the RoL. However, considering AI through a RoL lens is not only productive for the purposes of identifying concerns with AI systems; moreover, the RoL provides us with a framework and a lexicon that adds to - and complements - the existing debate on the use of AI systems and technologies, particularly in the context of decision-making in the public sector. There are promising developments across Europe. For example, within the CoE an Ad Hoc Committee on AI has been mandated in September 2019 to proceed with the development of a legally binding convention for the development, design and application of AI based on the CoE's standards on human rights, democracy, and the RoL.

The RoL is of course not the only lens through which we should approach regulation of these technologies; AI is a cross-sector, multi-disciplinary, un-bordered phenomenon that demands input from a range of disciplinary perspectives. But, the RoL has (perhaps characteristically, for the law) been slow off the mark has sat on the periphery of the AI discussion. In our view, it's time for the RoL to catch up with the debate and play a more prominent role moving forward. The scene has been set for the expansion of AI within the European Union, and the HLEG Guidelines and CoE Ad Hoc Committee on AI represent solid first-steps in solidifying the role of the RoL in moderating AI systems.

URL: https://binghamcentre.biicl.org/comments/71/catching-up-with-the-debate-artificial-intelligence-the-rule-of-law