Artificial Intelligence (AI) - Regulation and AI

1. Regulation **of** AI - "Keeping **us** safe from **them**"

- Arise in a context where we seriously entertained the possibility that an AI had achieved some degree of experiential awareness, such that its interest might conceivably matter to us
- Or at least, making amends when they harm us. Seen in proposals such as this from the EU Parliament around registers, kill-switches, and proposals for attribution of compensation liability when things go wrong
- E.g. Google Duplex - AI assistant - passed the Turing test (test to distinguish AI to a human)

2. Regulation **by** AI - Rule by algorithm

- What sorts of rules, or rule-making structures, should we have around the development or use of AI
- Example in NZ's society:
  - NZ supermarkets using facial recognition CCTV technology to identify shoplifters
  - ACC, Immigration NZ profiling algorithm
- As a result, Otago exports to work with Government on AI framework

3. Regulation **for** AI - "Keeping **them** safe from **us**"

- Addresses the question that arise when AI's themselves set or apply the rules
- Suggestions of what obligations we might owe to AI that become aware or are capable of suffering

Module structure

- The concept - what do we mean by AI, what do we mean by algorithms?
- The uses - in NZ but also other countries
- The problems - why would anyone be concerned about this?
- The solutions - what can we do about them? What solutions are put forward to address these concerns?
  - Solutions in NZ
  - Overseas solutions
- The problems with the solutions - because in the European context, there are pushback against some of the suggested checks and balances - either on the grounds that they won't make anything better, or perhaps even they would make things worse

The Concept

Regulatory challenge: *what* do we regulate? Definition of AI

- Michael Scherer: "Any AI regulatory regime must define what exactly it is that the regime regulates - in other words, it must define "artificial intelligence." Unfortunately, there does not yet appear to be any widely accepted definition of artificial intelligence even among experts in the field, must less a useful working definition for the purposes of regulation."
- HL Select Committee on AI (April 2018): "There is no widely accepted definition of artificial intelligence. Respondents and witnesses provided dozens of different definitions."

- **Discreetness** refers to the fact that AI development work can be conducted with limited visible infrastructure
- **Diffuseness** means that the individuals working on a single component of an AI system might be located far away from one another
- **Discreetness** refers to the fact that the separate components of an AI system could be designed in different places and at different times without any conscious coordination

- **Opacity** denotes the possibility that the inner workings of an AI system may be kept secret and may not be susceptible to reverse engineering.

Distinction - narrow or general (HL Select Committee)

- Artificial **general** intelligence refers to a machine with brand cognitive abilities, which is able to think, or at least simulate convincingly, all of the intellectual capacities of a human being, and potentially surpass them - it would essentially be intellectually indistinguishable from a human being
  - As of, doesn't exist yet
- **Narrow** AI systems perform specific tasks which would require intelligence in a human being, and may even surprise human abilities in these areas. However, such systems are limited in the range of tasks they can perform

Regulatory tension - the definitional problem

- Should regulators define broad or narrow?
- Should regulators be precise or flexible?

Regulatory challenge: *how* do we regulate?

- Do we need AI-specific rules? Or is it rather trying to make sure our general rules apply to the context of AI
  - Techno-exceptionalism - becoming fixated on the new technology and create laws specifically around that technology and thus lose sight of the fact that what regulators are concerned about are the same kinds of issues and problems that are already dealt with through existing law
    - Colin thinks there is a reasonable case to this, as what we are concerned with is discrimination, transparency, privacy, access to information that is held about us and control over the information = all issues that we are already concerned with in other settings = not new problems
  - Judge Frank Easterbrook and the Law of the horse
- Regulatory tilt (progress vs precaution) - what should the default setting for regulators be?

Algorithms

- Code or mechanism for translating some kinds of information into another kind of information, can take all kinds of forms
- Examples of predictive algorithms - taking Netflix history or Spotify history to recommend things you might like next, targeted advertising
- Increasingly being used by policy makers, government bodies like ACC, possibly even by landlords for tenants, banks, employers

---

The Uses

Various parts of government and crown agencies are using predictive analytics/ algorithms in NZ:

- Immigration NZ - algorithm used to profile overstayers
- IRD
- ACC
- MSD

We currently don't have an exhaustive list as to who are using it - but that might change soon

- Government to undertake urgent algorithm stocktake - jointly by Clare Curran and James Shaw (Minister for Government Digital Services and Statistics Minister)

AI Forum NZ

- Grouping of private sector businesses that area early uptakers of this technology
- Produced a report called "Artificial New Zealand Shaping a Future New Zealand" - was in part a survey of what is happening at the moment and a part a series of

projections about how they saw AI impacting on NZ society, economy, employment and various other aspects

- "In the economy, AI can be used to substitute human labour in a growing range of manual or repetitive tasks, enabling that same labour to be redeployed onto new, higher value tasks. Our modelling analysis finds that just through this labour conversion alone, AI has the potential to increase New Zealand GDP by up to $54 billion by 2035 across 18 industry classifications."
  - Degree of guesswork
- Very optimistic

Benefits of AI - Nuffield Council on Bioethics

- Medical applications of AI
- AI in the legal sector - better at menial tasks, such as:
  - Review documents and legal research
  - Help perform due diligence
  - Contract review and management
  - Predict legal outcomes
  - Automating divorce
- Once a certain type of document is denoted as relevant, machine learning algorithms can get to work to find other documents that are similarly relevant. Machines are much faster at sorting through documents than humans and can produce output and results that can be statistically validated. They can help reduce the load on the human workforce by forwarding on only documents that are questionable rather than requiring humans to review all documents

Automated weapons

- Better or worse?
- Particularly in occupied territories like Afghanistan and Iraq - incredibly mentally stressful for human soldiers, they crack and do terrible things

Use of algorithms within the criminal justice sphere

PredPol - The Predictive Policing Company

- In the USA, numerous police forces are using variants of this software
- Identifies crime hotspots
- Algorithm that uses historic data of crime detection, arrests, reports, and predicts a time, where the crime hotspots and turnouts are likely to be
- Predominantly used for deployment decisions
- In big cities, this is becoming quite granular - specifies the block or two
- Worrying ones are even more granular - some are purporting not just geographical areas but individuals they think are likely to commit a crime
  - Question then becomes: what do you do about that if you have that information available to you?
  - Is this moving into a pre-crime society? What can the police do?

Problems with predictive policing:

- Police already have an idea for deployment - so predictive deployment isn't new, now it is more specific
- It is self reinforcing - PredPol is never going to be wrong because if the police are dispatched to a particular area on a Saturday night to conduct stop and searches on people, they are going to find more low level crimes than if the police were not saturated in that area - then that data will be recorded in PredPol and locked in for future predictions that are now reinforced
- Or it could go the other way - if a particular area is flooded with police officers, there will be less of certain kinds of crime and the predictions are proven wrong by the very action they have taken in response to the prediction

COMPAS - Correctional Offender Management and Profiling Alternative Sanctions

- Used extensively through the US criminal justice system

- Used for assessment and case planning: everything from sentencing, what kind of correctional facility to be placed in, what kind of supervision should they have, parol conditions etc

The 'up' side to AI

- Predictions are more accurate - they exist anyway, might as well make it more accurate
- More efficient
- Remove human bias - AI doesn't have emotion

The Concerns

Why should we be worried?

1. Dignity and dehumanisation
2. Fettering of discretion
3. Transparency
4. Bias

Dignity and dehumanisation

Meg Leta Jones 'The right to a human in the loop: Political constructions of computer automation and personhood' (2017): "European political culture holds that treating an individual in a wholly automated way, or to provide only automated treatment, is to dehumanise the individual, because a machine can only treat a human in a computational manner. Thus, to treat a human in a wholly computational manner reduces the individual's dignity and restoration of dignity can be provided by a human in the loop.'

- Dehumanising decisions made by machines

Fettering of discretion

Cathy O'Neil: Weapons of Math Destruction

- Concerned about algorithms that make enormous decisions - decisions that affect vast numbers of people, or enormous quantities of money
- "You cannot appeal to a weapon of math destruction. That's part of their fearsome power. They do not listen. Nor do they bend. They're deaf not only to charm, threats, and cajoling but also to logic - even when there is good reason to question the data that feeds their conclusions'
- Whether that's a good thing or a bad thing will depend on a range of considerations. Discretion, after all, is the same mechanics that allows for nepotism and cronyism.
- She is claiming that even when you do have good reason to rebut the algorithm's response, a logical argument, evidence - the algorithm will say no

Transparency

If you believe you are subjected to a wrong decision made by an algorithm, it can only be challenged if you can understand on what basis the decision was made. You cannot challenge a decision that it was wrong if you don't know on what basis the decision was made.

John Edwards, Privacy Commissioner, 29 August 2016: "I encourage organisations to be transparent with their algorithms so that people can see why decisions were made - and challenge those decisions if they think they're wrong. I think that if you're going to make a judgement call about someone based on a data set, then that person has the right to see why you made that call. If you don't allow this transparency, and instead rely on the data without questioning it, then you could miss opportunities, or end up breaking laws against discrimination."

Why are algorithms not transparent?

1. Algorithms are proprietary

- Companies that sell them do not want to disclose their code to everybody else, they don't want to disclose precisely which input variables they use and what weightings they attach to those variables because if they did, people wouldn't have to pay them for it and construct their own
- Google are famously proprietal of their ranking algorithm that runs behind the Google search - unsure how it comes to the decisions and this is true for other examples of existing algorithms
- *Wisconsin v Loomis* (2015)
  - Loomis was subject to a decision made/informed to the extent by COMPAS software. Wasn't a decision about sentencing, was a decision about where he should serve his sentence - he wanted to be in a lower security prison near his family
  - Challenged the algorithm on various grounds:
    - He was not convinced that it was ever tested against a population like mine
    - Can't even see inside the algorithm - can't see how the decision is made
  - 'Northpointe, Inc., the developer of COMPAS, considers COMPAS a proprietary instrument and a trade secret. Accordingly it does not disclose how the risks are determined or how the factors are weighed.'
  - Purely legal question - can be resolved by legal means

2. Machine learning doesn't work that way

- Not amenable to that kind of transparency
- The Dark Secrets at the Heart of AI: "No one really knows how the most advanced algorithms do what they do. That could be a problem."
  - With regards to driverless cars: "As things stand now, it might be difficult to find out why. The system is so complicated that even the engineers who designed it may struggle to isolate the reason for any single action. And you can't ask it: there is no obvious way to design such a system so that it could always explain why it did what it did."

Why can't you design the system to explain their actions?

- Nature of more advanced forms of machine learning - 'deep neural networks'
- Algorithms are taught how to learn - teaching themselves to solve the problem better, learning from experience. Good in that the technology will continue improving
- Problem is that it is very difficult to know how they are now making the decisions - because the version of the software that was sold initially isn't exactly the same as the softwares a couple of years later as it has been learning, evolving and getting smarter over that time
- Algorithm used by NZ Corrections for decades ROC*ROI (risk of conviction, risk of offending) - that has to do with predicting future offending and future criminality
  - Very transparent algorithm - uses a small number of visible variables and the weightings on them
  - Not the way new algorithm is going to work
- Not just a legal question, couldn't even tell you even if you wanted to

Bias

The argument is that maybe algorithms will be less bias than humans because they don't play favourites. They don't have subconscious racial biases, gender biases, whatever it might be, they will do what they're told to do without bringing any emotions or cognitive biases into it.

Replicating/reinforcing bias

Attorney General Eric Holder, 2014: "Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualised and equal justice… they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."