**Artificial intelligence and law in New Zealand**

Principal investigators: Colin Gavaghan,¹ Ali Knott,² James Maclaurin.³

**Abstract**

A proposal for a three-year multi-disciplinary research project evaluating legal and policy implications of artificial intelligence (AI) for New Zealand. In addition to addressing general themes and challenges associated with AI, the project will contain several specific sub-themes, which will result in their own publications. It will conclude with a major publication drawing together the various topics addressed throughout its duration.

Major research questions will relate to:

• challenges posed by autonomous and semi-autonomous machines to notions of responsibility and culpability;
• questions of transparency and scrutability posed by decision-making algorithms, with particular reference to the criminal justice system;
• questions relating to employment, and in particular the replacement of human jobs by AIs;
• questions of 'machine morality', and the normative rules with which autonomous machines will be programmed.

Methodology will include:

• extensive review of the large and rapidly expanding body of literature on AI and law;
• consultation with international experts, in part through a series of multi-disciplinary workshops;

---

¹ Associate Professor, Faculty of Law, University of Otago.
² Associate Professor, Department of Computer Science, University of Otago.
³ Associate Professor, Department of Philosophy, University of Otago.
examination of practice in other jurisdictions, in particular the USA, and interviews with stakeholders there, with a view to informing discussion of policy options for New Zealand.

Contents of proposal

Introduction 3
Project structure 5
Justification for expenditure 7
General research themes 10
Theme One: Predictive algorithmic analysis and the New Zealand criminal justice system 12
Theme Two: Artificial Intelligence, Employment and Law (Near-future AI technologies; regulation ‘of’ AI) 19
Proposed outputs and dissemination plans 26
Proposed timeline 26
Budget 27
Introduction

Both in New Zealand and internationally, a consensus is building that the time is right for research into the social, ethical and legal implications of artificial intelligence.

In launching its inquiry into artificial intelligence and robotics in March 2016, the UK Parliament’s Science and Technology Committee expressed the hope that ‘it will be soon enough to be productive and late enough to be relevant.’ 4 In October 2016, the US President’s National Science and Technology Council published Preparing for the Future of Artificial Intelligence, a substantial part of which focused on regulation of AI. 5

Perhaps most famously, since 2015, thousands of prominent scientists and entrepreneurs (including Stephen Hawking, Steve Wozniack and Elon Musk) have put their names to an open letter, calling for ‘expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial’, including contributions from ‘economics, law and philosophy to computer security, formal methods and, of course, various branches of AI itself.’ 6

In New Zealand too, there has been recent recognition of a need for attention to be turned to this question. In October, the Foreword to a joint report by the Institute of Directors and Chapman Tripp concluded that: 7

AI will raise major social, ethical, and policy issues in almost every sector.

It is critical – for New Zealand’s sake – that we actively consider, lift

---

4 Science and Technology Committee, Report: Robotics and artificial intelligence (2015), at [8]
5 US President’s National Science and Technology Council, Preparing for the Future of Artificial Intelligence (October 2016) https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
7 S Arcus, Chief Executive Institute of Directors; B McClintock, Head of Technology, Media and Telecommunications Chapman Tripp. Foreword to Determining our future: Artificial Intelligence. Opportunities and challenges for New Zealand: A call to action. (October 2016)
awareness of, and prepare for the changes AI will bring. This work needs to start now.

We agree that the time is also right for this research to begin in earnest in New Zealand. Throughout 2016, a group of researchers from across Otago University have been meeting regularly to discuss the technological, social, ethical, economic and legal implications of AI. Initiated by Ali Knott (Computer Sciences) and James Maclaurin (Philosophy), the AI and Society Group has also drawn on expertise from politics, economics, psychology and law, and is already liaising with internationally recognised experts.

The Group has spent the initial year of its existence sharing expertise and sources, via a series of presentations and its wiki page, and identifying questions for future investigation. It is now ideally positioned to undertake a major project of original research.

In this proposal, we set out a research strategy that will begin to address the need identified in the IoD report. Importantly, we also believe this proposal aligns well with several of the key themes identified by the NZ Law Foundation within its ILAPP project. Most specifically, we believe it addresses the need identified in the following paragraph:

The impact of artificial intelligence (AI) also needs to be looked at as the impact of this is expected to be huge in the short term. How could it affect the outcome of court decisions, and law generally, including the practice of law, as well as the impact of AI on wider society?

As well as answering specific questions, we believe the project will allow us to build capacity within New Zealand to address the as yet unforeseen questions that will be posed by future developments in AI.

---

8 https://ai-and-society.wiki.otago.ac.nz/index.php/Main_Page
**Project structure**

Our proposed project spans the three year duration of the ILAPP project. To accommodate the broad range of research questions that could be addressed, we will focus on two separate themes: Theme One targets an existing application of AI, while Theme Two targets an application on the near horizon. These sub-projects raise a mixture of immediate and intermediate legal and societal questions. We will produce publications on each of the two application areas. However, there are many common themes running across the two areas. At the project’s end we will also produce a major publication reflecting more generally on the legal and social implications of AI technologies.

The major output will take the form of two major publications which will ultimately draw together all of our major findings and recommendations. These will be published at the end of the second and third years of the project respectively.

We have chosen our two themes to reflect two fundamentally distinct ways in which legal/regulatory issues bear on AI. Drawing on the field-leading work of Roger Brownsword, it would distinguish between Artificial Intelligence as a regulatory tool, and as a regulatory target; that is, regulation *of* AI, and regulation *by* AI. Our first theme focuses on regulation *by* AI; our second theme focuses on regulation *of* AI.

Regulation *of* AI may relate both to immediate/certain and more distant/speculative aspects. Some of these – such as the paradigm case of driverless cars - may be amendable to New Zealand-specific solutions, while others may depend more on international collaboration. Even in the latter cases, however, we consider that New Zealand could serve as a model of how such

---

technologies could be approached.

Regulation by AI has thus far received less attention both academic and media. However, the use of machine learning techniques is becoming widespread in policing, and increasingly in other aspects of the criminal justice system. Indeed, the use of risk predictive software in informing sentencing decisions has recently been the subject of major litigation in the USA.

The first part of this proposal provides an outline of the sorts of general themes and challenges that we anticipate addressing throughout this project. The second and third parts specify two particular contexts in which the emergence of AI poses potentially novel legal and regulatory questions.

Cohesion: the multi-disciplinary nature of our team will be invaluable in allowing us to approach these issues from a variety of perspectives.

Colin Gavaghan’s expertise relates to the relationship between law and emerging technologies. He has extensive experience of research and publication relating to the regulation of reproductive, genetic, neuro- and other technologies. His role will involve research into possible legal and regulatory challenges and strategies, and comparative analysis of international legal developments.

Ali Knott’s expertise is in machine intelligence. His role will be invaluable in evaluating the real possibilities for AI, explaining the manner of its operation, and scrutinising the claims made by its more enthusiastic proponents.

James Maclaurin’s expertise is in the philosophy of science. His role will involve analysis of ethical and philosophical concepts, including ontological issues in characterising artificial intelligence and robotics, the ‘moral rules’ with which machines may be programmed, questions of responsibility and accountability, and questions of fairness.
Justification for expenditure

1. Postdoctoral researchers. The largest expenditure in our bid relates to the funding of postdoctoral research assistance. Our proposal is to hire two postdocs, each for two years. Based on Otago’s Research Office projected salary scales, they will be paid as follows:

2017: $77,037
2018: $78,578
2019: $80,150

Each postdoc will be allocated to one of the major themes, and employed for a period of two years. The middle year of the project will see both postdocs employed together.

While the principal investigators will themselves by active participants in the research, the scale of the existing and emerging literature in this field, and the rapid pace of developments both in the technology and the legal responses thereto, mean that assistance in conducting a thorough literature review will be necessary.

It is our belief that the volume and complexity of the literature in this area require researchers of an advanced level, i.e. those capable not only of identifying potentially relevant material, but of understanding it. This will allow them to filter, organise and summarise that material in a form that will genuinely assist the principal researchers. In addition to being novel, the project is highly interdisciplinary in nature. Since it is highly unlikely that we will be able to locate a researcher who is qualified in law, computer science and philosophy, we will require someone who is capable of ‘upskilling’ in disciplines other than their own. This, we believe, will require mature researchers, with a high capacity for autonomous working. They should already have a PhD.
While a less qualified researcher would cost less, we believe that doing justice to this subject necessitates high-level assistance.

2. Travel. While the technologies we will be researching have not yet arrived in New Zealand to any significant extent, several of them are already being utilised in other jurisdictions. In our view, proper preparation for their arrival in New Zealand will require first-hand examination of the manner of their use in those places where they are being used.

We therefore propose a series of visits to jurisdictions where the technologies we will be examining are already in use. While it is difficult to be precise until the research begins in earnest, in relation to Theme One, we would anticipate that this is most likely to be the United States, where predictive policing and sentencing software is already being widely used.

We will attempt to meet and interview end users and stakeholders with direct experience of these systems, for example, representatives of relevant police forces and correctional bodies who have experience of using predictive technologies, as well as civil rights organisations who have expressed concerns as to that use.

Similarly, with regard to Theme Two, we hope to have a chance to meet with and interview ‘early adopters’ of ‘AI employees’, including possibly law firms. Again, the United States seems the most promising location for such research.

3. Workshops
The project will also involve two workshops. These will bring together researchers and stakeholders from within New Zealand – together with a few high-level international guests – to present on, discuss and critique approaches to the themes of our research. It is our intention that these workshops should directly inform the major published outputs from the project.
We have budgeted for a total of 16 international flights, at an estimated $2800 per flight. These will be allocated between our fact-finding trips to other jurisdictions, and the international guests we bring to New Zealand. In terms of accommodation, we have budgeted for an average travel duration of 14 nights at $170 per night.

4. Project co-ordinator
The project will require organisation and co-ordination – for example, booking flights and accommodation, for the PIs and for guests; arranging meetings with overseas experts and stakeholders; co-ordinating workshops. The time each of us has to spend in this project is finite (particularly in view of James Maclaurin's role as associate dean and Ali Knott's private sector work with Soul Machines), and it would be a considerable asset for us to be freed up to focus on research and writing.

We are also keen that the burden of these tasks not be shifted to the administrative staff at our respective departments.

We have therefore budgeted for the employment of a project co-ordinator who will undertake these tasks. We have proposed that this be Research Assistant Level 3, for 8 hours per week for the life of the project.

5. Publication costs
While it is our intention that our major outputs will be published via a recognised academic publisher, we also intend to produce and disseminate a brief report, summarising our main findings, to be distributed among key stakeholders in New Zealand. This item is to cover the publication and distribution costs.
General research themes

Running across both our chosen themes are various general issues regarding regulatory challenges and strategies pertaining to AI. Identifying these will be a key goal of the project, but they are likely to include the following:

1. Whether an AI-specific regulatory response is merited, or whether existing laws and norms could (perhaps with some modification) be applied? The US National Science and Technology Council proposed that\(^\text{10}\)

   If a risk falls within the bounds of an existing regulatory regime, moreover, the policy discussion should start by considering whether the existing regulations already adequately address the risk, or whether they need to be adapted to the addition of AI.

   The challenge, of course, will be to determine whether the existing regulatory regime does indeed adequately address any additional risks posed by the use of AI.

2. If an AI-specific response is merited, how for these purposes is ‘AI’ to be defined and recognised? A recent report in *Nature* is just the latest to acknowledge that ‘there is no consensus on what counts as AI.’\(^\text{11}\) The remit of any AI-focused law or regulatory body may depend on the ability to define their subject matter with sufficient precision.

3. Questions of regulatory phase (when should such strategies be applied?) and tilt (what should be the default setting of the regulator? Where, for example, should the balance be struck between progress and precaution, or between liberty and safety?)\(^\text{12}\)

---

\(^{10}\) *Preparing for the Future of Artificial Intelligence*, p1.

\(^{11}\) K Crawford and R Calo ‘There is a blind spot in AI research’, *Nature*, 13 October 2016.

\(^{12}\) These concepts are derived from the influential work of Roger Brownsword. See for example R Brownsword *Rights, Regulation, and the Technological Revolution* (Oxford: Oxford University
4. Human-machine interfaces. Many applications of AI are likely to involve the technology operating in tandem with a human operator. For example, it is likely that ‘self-driving cars’ will initially require a human to be present and prepared to assume control if required. Doubts exist, however, about human capacity to remain inactive but alert for prolonged periods, and about the likelihood that we will unthinkingly defer to the machine.

5. The issue of normative algorithms and ‘machine morality’ has recently been brought to the fore by Mercedes’ much publicised announcement that their self-driving cars will be programmed to prioritise the safety of their passengers rather than other drivers or pedestrians. While ‘machine learning’ will allow AIs to develop in unpredictable ways and develop novel solutions, the question of programming them with moral rules from which they cannot depart may offer reassurance or concern.

It is proposed that this part of the project would involve liaison and collaboration with Michael Cameron’s research project on driverless cars. Discussions have already taken place with a view to ensuring that the research teams will complement, and not duplicate, one another’s efforts.

Expected collaborators to include:

- Russell Blackford (University of Newcastle (Aus), law and philosophy). Co-editor of *Intelligence Unbound: The Future of Uploaded and Machine Minds*
- Dr Joanna Bryson (Department of Computer Science, University of Bath). Expert in artificial intelligence.
- Michael Cameron (Lead legislation solicitor, Department of Corrections). Researcher in legal and regulatory issues around driverless cars.
- John Danaher (University of Galway, law) Expertise in emerging technologies generally, and artificial intelligence particularly.

• James Every-Palmer (Barrister, Stout Street Chambers) Qualifications in artificial intelligence, with particular interest in applications to law.
• Associate Professor Brendan McCane (Otago, Comp Sci) Expertise in AI.
• Professor David O’Hare (Otago, Psychology) Expertise in cognitive engineering and human decision making.
• Sean Goltz (Waikato, law) Expertise in law and emerging technologies.
• Professor Rob Sparrow (Monash University, philosophy) Applied ethics and philosophy, with particular interest in AI and robotics.

**Theme One: Predictive algorithmic analysis and the New Zealand criminal justice system (Current AI technologies; regulation ‘by’ AI)**

In a 2014 report on preventive detention in New Zealand,\(^{13}\) funded by the Law Foundation, researchers at Otago University noted a general trend away from a backward-looking, reactive/retributive model of criminal justice, and towards a forward-looking, predictive/preventive model. This trend has also been observed in other jurisdictions. As Professor Bert-Jaap Koops of the Tilburg Institute of Law and Technology has expressed it:\(^{14}\)

society’s architecture is slowly being changed ... Not only to prevent crime with real and metaphorical hinges and locks ... but also to better enable the detection of future crimes being planned or committed. This is a fundamental change.

This shift has manifested itself in a variety of forms. The 2014 Otago report focused on measures implemented following criminal conviction; forms of criminal and civil detention oriented explicitly towards the avoidance of future offending. The preventive shift, however, is also manifested in policing practices, which are increasingly coming to be supported by actuarial and predictive

---
\(^{13}\) C Gavaghan, J Snelling and J McMillan. *Better and Better and Better? A Legal and Ethical Analysis of preventive detention in New Zealand* (November 2014)

techniques and technologies. A recent report from the USA described the PredPol software package, ‘a startup predictive policing vendor [that] applies machine learning techniques to try to predict future crimes at locations other than where previous crimes took place.’\textsuperscript{15} It has recently been reported that PredPol is being used by 60 police departments throughout the US.\textsuperscript{16}

While PredPol’s use has thus far largely been limited to informing decisions about police deployment, other software developers are making more ambitious claims. The “Strategic Subjects List” deployed in Chicago prepares a “heat list” ‘of people that a computer program has judged most likely to become involved in a shooting, either as a perpetrator or as a victim.’\textsuperscript{17} As we discuss below, this has raised various concerns, including the possibility that it might mean that ‘our historic biases and prejudices are being reified in machines.’\textsuperscript{18}

If the use of predictive algorithms is likely to be controversial in policing, it is proving no less contentious in other parts of the criminal justice system. Risk-assessment software is now widely employed in various aspects of the post-arrest and post-conviction process in the USA. One of the most commonly used, Northpointe’s COMPAS software, “has modules designed for pretrial, jail, probation, prison, parole and community corrections applications.”\textsuperscript{19}

The advantages of such technologies should not be overlooked. There are clearly positives to police resources being deployed more efficiently, and the prospect of criminal activity being prevented rather than detected has evident appeal. Similarly, insofar as decisions about bail, sentencing, parole should be informed by assessments of future risk, those assessments should be as accurate as

\footnotesize
\begin{enumerate}
\item Upturn, \textit{Stuck in a Pattern: Early evidence on "predictive policing" and civil rights} (August 2016)
\item ‘Can “predictive policing” prevent crime before it happens?’ \textit{Science}, 28 September 2016
\item Upturn, \textit{Stuck in a Pattern: Early evidence on "predictive policing" and civil rights} (August 2016)
\item A Caliskan-Islam, J J Bryson, A Narayanan. ‘Semantics derived automatically from language corpora necessarily contain human biases’ arXiv:1608.07187 (August 2016)
\item J Angwin, J Larson, S Mattu and L Kirchner. “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks.” \textit{ProPublica}, May 23, 2016.
\end{enumerate}
possible.

The enthusiasm for such techniques also, however, gives several reasons for caution, even concern. Although some early indications have been quite promising, there is to date a lack of reliable data as to the efficacy of such programmes. Furthermore, concerns have been expressed about what Evgeny Morozov has called ‘the perennial problem of algorithms: their presumed objectivity and quite real lack of transparency.’ The latter issue – the inability to scrutinise the basis on which the programmes make their predictions – is potentially exacerbated by commercial sensitivity.

The Upturn report noted that ‘vendors routinely claim that the inner working of their technology is proprietary, keeping their methods a closely-held trade secret, even from the departments themselves.’ In the context of risk-assessment software in sentencing, a ProPublica report recently claimed that Defendants rarely have an opportunity to challenge their assessments. The results are usually shared with the defendant’s attorney, but the calculations that transformed the underlying data into a score are rarely revealed.

In a recent challenge to the use of the COMPAS technology in a sentencing decision, the Supreme Court of Wisconsin held that, in order to comply with due process requirements, the use of COMPAS must be accompanied by a warning to the sentencing judge that ‘the proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined.’

---


21 Upturn, *Stuck in a Pattern: Early evidence on “predictive policing” and civil rights* (August 2016)


23 *Wisconsin v Loomis* 2016 WI 68, at ¶66.
This is one of the distinct and novel challenges of the current generation of AI algorithms: whether or not their judgements are correct, or contain particular biases, there is a real possibility that their human users have no way of understanding the basis on which these judgements are made. If a machine cannot justify its judgements in a way that its human users can understand, or the justifications it reports inadequately reflect its computational processes, then their acceptance of these judgements would appear ultimately to be a matter of faith in authority. Does this raise problems in principle with the use of such systems in making serious life-affecting decisions?

It is also important to remember that predictive technologies validated for use in one environment may not be sensitive to differences on others.

Concern has also been expressed about the putative objectivity of such systems. As Morozov has asked:24

how do we know that the algorithms used for prediction do not reflect the biases of their authors? ... Might algorithms – with their presumed objectivity – sanction even greater racial profiling?

Such concerns are given additional credence by the findings of the ProPublica report into the Northpointe programme:

The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.

Questions also arise in relation to the decisions that would be informed by such algorithms. While police deployment decisions may be relatively uncontentious the same cannot be said for more invasive practices such as ‘stop and search’. Decisions about sentencing, parole and preventive detention, predictably, would be most contentious of all.

---

This project aims to evaluate the prospects for predictive technologies in a New Zealand policing and criminal justice context. It will adopt a multi-disciplinary approach, drawing on expertise from law, computer science and philosophy. Our key research questions will include:

- What options are available in terms of risk/recidivism predictive programmes? Which are in use, and what is known about their efficacy? What is known about their future development? (We anticipate drawing heavily on US sources here, as – from our initial investigations – this is where the most extensive, or at least extensively documented, use is being made.)
- To what extent are police, courts or other participants in the New Zealand criminal justice system using predictive software, or considering doing so in future?
- What use is being made of the results generated by such systems? To what extent is human oversight still a factor? In July, the Supreme Court of Wisconsin sanctioned the use of the COMPAS risk assessment algorithm, provided judges receive written warnings about the value of such scores and do not rely on them exclusively. But is this realistic? Psychology research suggests that, when operating a dual decision making system like this, humans tend to over-rely upon or defer to the machine component.
- What means have been used to validate the accuracy of these systems, and in particular, to validate them for use on a New Zealand population?
- What capacity exists for the results generated by such systems to be scrutinised and challenged?
- What unintended knock-on legal effects may be generated by the use of such software? (For example, in the USA, the designation of a geographical

---

25 Wisconsin v Loomis 2016 WI 68
26 In Loomis, the Supreme Court of Wisconsin required that COMPAS technology must be subjected to ‘certain cautions’, including that it ‘must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.’ At ¶66.
locale as ‘high risk’ has been held by the Supreme Court to lower the threshold for ‘probable cause.’)

- What social, political and ethical considerations should we be taking into account before deploying such systems, or basing other decisions upon their results?

Expected collaborators to include:

- John Danaher (University of Galway, law)
- Tom Douglas (Oxford University, Uehiro Centre, ethics) Senior Research Fellow in the Oxford Uehiro Centre for Practical Ethics. Expertise in technologies of crime prevention.
- Geoff Hall (University of Otago, law) Professor Hall is New Zealand’s foremost academic authority on sentencing law.
- Bernadette McSherry (University of Melbourne, law) Foundation Director of the Melbourne Social Equity Institute at the University of Melbourne. Internationally recognised authority, and author of numerous books on post-sentence detention and the role of risk prediction in the criminal justice system.

**Methodology and budget**

This part of the project will require the recruitment of a postdoctoral researcher, whose role will involve a mixture of technical research and research into how technology is applied. While the principal investigators will of course themselves be conducting original research, the anticipated scale of this project will render such assistance invaluable.

The researcher will have to learn (at a high level) how predictive policing systems work. Ali Knott will provide support for this role. It is not anticipated that the researcher will need to be a computer scientist, but s/he should be sufficiently computer literate to enable them to learn something about the technology.
More specifically, it is anticipated that the researcher will conduct a literature review on the topic of predictive policing, identifying current trends in the US, which is at the forefront of applications in this area. There will be an emphasis on understanding (i) what the technologies do - what data they use, how they learn; (ii) how the technologies are applied.

The second phase of the research on this theme will involve visits to selected jurisdictions where different forms of predictive technologies have been utilised, with a view to reviewing first hand how they are used. It is anticipated that these jurisdictions will be in the United States. We will attempt to meet and interview, for example, with representatives of relevant police forces and correctional bodies who have experience of using predictive technologies, as well as civil rights organisations who have expressed concerns as to that use. Relevant ethics approval will be obtained prior to all such interviews.

The third phase will involve discussions with New Zealand stakeholders, including the NZ Police and the Department of Corrections. This will be done with a view to establishing whether there are plans in place to begin utilising such technologies, and to sharing our findings and recommendations. Again, relevant ethics approval will be obtained prior to all interviews.
**Theme Two: Artificial Intelligence, Employment and Law (Near-future AI technologies; regulation ‘of’ AI)**

Technological unemployment is a type of structural unemployment due to the replacement of workers by machines.\(^\text{27}\) It is, of course, not a new phenomenon, but fears about mass technological unemployment have recently become widespread in public debate, in a wide range or public policy settings, and in academic research. This concern is driven primarily by the development of artificial intelligence that threatens to replace humans in many highly skilled, high status occupations. It is now documented in a large number of popular academic books typified by Erik Brynjolfsson and Andrew McAfee’s *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*.

In their much-discussed report,\(^\text{28}\) Carl Frey and Michael Osborne argue that that over 47 percent of current jobs will likely be automated in the next two decades. Even more ominously, Benzell *et al* argue that artificially intelligent machines will eventually replace humans in the workplace just as a variety of agricultural machinery has virtually replaced horses on farms.\(^\text{29}\) While we have concerns about the methods of some recent studies, we think there is sufficient evidence to warrant urgent consideration of the legal implications of a near-to-medium-term future in which technology radically alters the world of work. Such implications might include the following.

**Theme 2a. Predictive programming and employment**

While Theme 1 considered predictive programming in the context of the criminal justice system, it is also increasing relevant in decisions around of employment. AI-informed hiring decisions may give the impression of being immune from


discrimination. Yet, in the same manner as its uses in the CJS, it may be that such systems are merely masking discriminatory practices, cloaking them in a veneer of objectivity, while in fact reflecting and entrenching problematic assumptions:

AI hiring systems have the troubling potential to encode, and perhaps amplify existing biases against historically disadvantaged groups ... correcting bias and protecting workers does not happen without keen attention, and sometimes a metric needs to be discarded altogether. Such systems must be closely examined and audited to assess disparate impact, analyzing who is getting hired, who is not, and why.30

Again, as with other uses of AI, questions of scrutiny and accountability may arise concerning decisions to hire, fire, promote and – increasingly in the age of ‘precarious’ employment – allocate shifts. As the artificialintelligencenow report points out, it may be that such decisions will increasingly come to be seen not as management decisions at all, but rather, as the outputs of objective automated processes, and hence as effectively immune to challenge.

Question: how can algorithmic employment decisions be subjected to the kinds of scrutiny and accountability that is applied when such decisions are made by human managers?

**Theme 2b. Dismissal and redundancy**

At present, New Zealand employment law relies on an important distinction between redundancy and unfair dismissal. An employee whose role becomes genuinely superfluous can rely only on limited legal protections (subject to any contractual obligations or collective agreement), relative to one who is replaced by another employee. Thus far, employment law has had no difficult in conceiving of jobs replaced by mechanisation as falling within the former category. But as AI

---

30 The Social & Economic Implications of Artificial Intelligence Technologies in the Near-Term July 7th, 2016; New York, NY.
https://artificialintelligencenow.com/media/documents/AI_NOW_LABOR_PRIMER.pdf
technology comes to assume rather than displace formerly human roles, questions arise as to the whether this is the correct conception of what is taking place. A human worker replaced by a machine is one thing, but when law firms are spoken of as hiring “AI lawyers”, we may wonder whether that assumption is adequate.

Questions: how will the displacement of human workers with AI fit within existing categories of redundancy and unfair dismissal? Does the ‘AI revolution’ present a novel challenge for employment law in this regard, and if so, how might it be met? Could a case be made for legal protections in favour of human workers, perhaps along the lines of measures that allow the hiring of migrant labour only where the domestic workforce is unable to meet the demand?

Theme 2c. Mechanisation and the professions

While the prospect of technological redundancy is far from new, one unusual aspect of the perceived threat from AI is its implications for highly skilled professional jobs, a section of the economy that has appeared relatively insulated from past ‘technological revolutions.’ As the partnership between major US law form BakerHostetler and AI firm ROSS Intelligence demonstrates, the legal profession is unlikely to be immune to such implications.

The possible advantages of such an approach are easy to see. ‘AI lawyers’, it is claimed, can often trawl data and deliver answers quicker and cheaper than their human counterparts – and in time, perhaps, with greater accuracy. The disadvantages, most obviously, could be felt by human lawyers. It is possible, however, that other less evident losses may be experienced at a societal level as greater number of professional roles are taken over by AI systems.

---

31 The world’s first artificially intelligent lawyer was just hired at a law firm. Business Insider UK, 16 May 2016.
In addition to the obvious aspects of their roles, professions such as lawyers, doctors and teachers have important legal, social and ethical responsibilities that may be seen as integral to those professions.

- The Medical Council of New Zealand’s Good Medical Practice (2016), for example, sets out its guide to clinical competence, cultural competence and ethical conduct for doctors as required by law.\(^{32}\) The guide requires *inter alia* that doctors ‘demonstrate reflectiveness, personal awareness, the ability to seek and respond constructively to feedback and the willingness to share knowledge and to learn from others’ and ‘balance their duty of care to each patient with your duty of care to the community and wider population’;

- Section 4 (a) (v) of the Education Act 1989 requires that universities accept the role “as critic and conscience of society”.

- New Zealand lawyers are subject to a statutory ‘obligation to uphold the rule of law and to facilitate the administration of justice in New Zealand’.\(^{33}\)

The type of machine intelligence that we propose investigating—dialogue systems leveraging machine learning on large datasets—might be able to do the conveyancing on your house, diagnose your gastroenteritis or teach your children calculus. The concern, however, is that the gradual implementation of such technologies might erode crucial social functions provided by human lawyers, doctors, and teachers in a way that would be difficult to ameliorate. A small number of people employed to act as critic and conscience of society could not replace the function currently performed by the large cohort of academics embodying, as they do, a very broad array of specialist knowledge.

Questions: does the replacement of professionals such as lawyers, doctors and teachers with AI alternatives risk eroding something of social and ethical value in the professions? If so, how might this be safeguarded?

---

\(^{32}\) Section 118(i) of the Health Practitioners Competence Assurance Act 2003.

\(^{33}\) Section 4(a), Lawyers and Conveyancers Act 2006.
**Theme 2d. The prospect of electronic persons?**

In a 2015 draft resolution to the European Parliament, the European Union’s Committee on Legal Affairs called on the European Commission to consider:\(^{34}\)

creating a specific legal status for robots, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations, including that of making good any damage they may cause, and applying electronic personality to cases where robots make smart autonomous decisions or otherwise interact with third parties independently.

The attribution of legal personality to non-humans is not without precedent. In this context, such a device may have the potential to answer or ameliorate some of the concerns presented by AI in the workforce, including those related to liability, redundancy and the erosion of important professional functions. Such mechanisms also have potential to decrease the differential between the costs of human and robotic workers and hence, to be a further means of protecting jobs in general or particular occupations in which we wish to discourage the displacement of people by robotic workers.

It is also worth considering the extent to which paid employment presently provides the mechanism whereby collective risk-pooling systems such as national insurance and, in the New Zealand context, ACC are funded. Another suggested advantage of the attribution of electronic personhood is that it would preserve intact the models of employer and state contribution whereby such funds are maintained.

On the other hand, proposals to treat machines as persons are fraught with technical, legal and philosophical problems. Terms such as “robot” and “artificial

---

\(^{34}\) European Parliament Committee on Legal Affairs, MOTION FOR A EUROPEAN PARLIAMENT RESOLUTION with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))
intelligence” are ontologically vague. Treating every piece of machinery that might potentially displace a human worker as a robot would have the implausible consequence of extending personhood to productivity-enhancing tools in general. Even very simple machines such as self-opening doors are capable of limited autonomy based on basic ‘sensory’ and information processing functions. They are, in this respect, like very simple organisms. But where on the on the scale from simple stimulus-response systems to fully fledged human level machine intelligence, do we mark the bounds of artificial intelligence for legal purposes?

Questions: is ‘electronic personhood’ a possible – even partial – solution to some of the challenges posed by an artificially intelligent workforce? What problems would such an approach face?

Expected collaborators to include:

- John Danaher (University of Galway, law)
- Professor Rob Sparrow (Monash University, philosophy) – Applied ethics and philosophy, with particular interest in AI and robotics.
- Associate Professor Lisa Ellis (Otago, Philosophy, Politics and Economics) – interest in ethical and economic implications of robotics and AI, especially for the workplace.
- Wayne Rumbles (Waikato, law)
- James Every-Palmer (Barrister, Stout Street Chambers) – AI
- Carol Jess – lecturer and PhD candidate, Centre for Labour, Employment and Work, VUW. Expertise in theories of labour and future of work.

**Methodology and budget (Theme 2)**

This part of the project will require the recruitment of a postdoctoral researcher, who will ideally have a background in law, but (again) be computer-literate and have interests/aptitude in philosophical issues.

For part (a) ‘Predictive programming and employment’, the methodology would approximate that for Theme 1, that is, (i) a literature review of the current generation of hiring systems and their application, then (ii) discussion with some
selected companies about the use of these systems; then (iii) an analysis of how the systems' recommendations can be justified/scrutinised. This work should be done in collaboration with the RA for Theme 1.

For part (b) 'Redundancy and dismissal', the RA could perhaps start with an analysis of the New Zealand laws covering these two areas, and an analysis of laws (in New Zealand and elsewhere) favouring nationals over foreigners. These both seem to require a legal background. But the RA would also have to consider the wider question of how to define an AI system that is sufficiently like a person for the mooted laws to kick in: that's a proper mixture of philosophy and technical AI issues. (This question of how to define AI in legislation shows up in several other parts of this theme - it's a central question for part (d).)

For part (c) 'Mechanisation and the professions' it might be helpful to come up with a taxonomy of ways in which human members of the professions (academics, lawyers, doctors) contribute to society over and above the content of their jobs. Then the RA could consider whether there are differences between these social contributions, and whether legislation should treat them differently.
**Proposed outputs and dissemination plans**

- Two edited collections, on the model of the Otago Genomics Project. Our intention, however, would be to publish through a major academic publisher.
- Report of key findings, to be distributed among key New Zealand stakeholders.
- 3-4 peer-reviewed articles or chapters.
- Two workshops.
- Conference presentations.
- Media.
- Website.

**Proposed timeline**

**Year 1**
Employment of postdoctoral researcher 1
Focus on general issues and Theme 1

**Year 2**
Employment of postdoctoral researchers 1 and 2
Workshop 1
Publications relating to Theme 1
Commence research on Theme 2

**Year 3**
Employment of postdoctoral researcher 2
Workshop 2
Publications relating to Theme 2
# Budget

<table>
<thead>
<tr>
<th>Funding category</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Staffing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postdoc (Theme 1)</td>
<td>82638</td>
<td>84291</td>
<td></td>
</tr>
<tr>
<td>Postdoc (Theme 2)</td>
<td></td>
<td>84291</td>
<td>85997</td>
</tr>
<tr>
<td>Project coordinator</td>
<td>11001</td>
<td>11221</td>
<td>11445</td>
</tr>
<tr>
<td><strong>Travel and Accommodation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airfares International</td>
<td>14933</td>
<td>14933</td>
<td>14933</td>
</tr>
<tr>
<td>Accommodation</td>
<td>12693</td>
<td>12693</td>
<td>12693</td>
</tr>
<tr>
<td>Taxis and car rentals</td>
<td>1066</td>
<td>1066</td>
<td>1067</td>
</tr>
<tr>
<td>Travel Insurance etc.</td>
<td>2846</td>
<td>2846</td>
<td>2846</td>
</tr>
<tr>
<td><strong>Publication costs</strong></td>
<td>0</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td><strong>Equipment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laptops / software</td>
<td>4000</td>
<td>4000</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>129177</td>
<td>220341</td>
<td>133981</td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td></td>
<td></td>
<td>483499</td>
</tr>
</tbody>
</table>