

## Ethics and Regulation of Artificial Intelligence

Anthony Wong

### ▶ To cite this version:

Anthony Wong. Ethics and Regulation of Artificial Intelligence. 8th IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KM), Jan 2021, Yokohama, Japan. pp.1-18, 10.1007/978-3-030-80847-1\_1 . hal-04041349

# HAL Id: hal-04041349 https://inria.hal.science/hal-04041349v1

Submitted on 22 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

## **Ethics and Regulation of Artificial Intelligence**

Anthony Wong<sup>1,2,3</sup>

<sup>1</sup>Managing Director, AGW Lawyers & Consultants, Sydney, Australia 
<sup>2</sup>Vice-President, IFIP 
<sup>3</sup>Past President, Australian Computer Society (ACS) 
anthonywong@agwconsult.com

Abstract. Over the last few years, the world has deliberated and developed numerous ethical principles and frameworks. It is the general opinion that the time has arrived to move from principles and to operationalize on the ethical practice of AI. It is now recognized that principles and standards can play a universal harmonizing role for the development of AI-related legal norms across the globe. However, how do we translate and embrace these articulated values, principles and actions to guide Nation States around the world to formulate their regulatory systems, policies or other legal instruments regarding AI? Our regulatory systems have attempted to keep abreast of new technologies by recalibrating and adapting our regulatory frameworks to provide for new opportunities and risks, to confer rights and duties, safety and liability frameworks, and to ensure legal certainty for businesses. These past adaptations have been reactive and sometimes piecemeal, often with artificial delineation on rights and responsibilities and with unintended flow-on consequences. Previously, technologies have been deployed more like tools, but as autonomy and self-learning capabilities increase, robots and intelligent AI systems will feel less and less like machines and tools. There is now a significant difference, because machine learning AI systems have the ability 'to learn', adapt their performances and 'make decisions' from data and 'life experiences'. This paper presented at the International Joint Conference on Artificial Intelligence - Pacific Rim International Conference on Artificial Intelligence in 2021 provides brief insights on some selected topical developments in ethical principles and frameworks, our regulatory systems and the current debates on some of the risks and challenges from the use and actions of AI, autonomous and intelligent systems. [1]

**Keywords:** AI, Robots, Automation, Regulation, Ethics, Law, Liability, Transparency, Explainability, Data Protection, Privacy, Legal Personhood, Job Transition, Employment.

#### 1 Introduction

AI and algorithmic decision-making will over time bring significant benefits to many areas of human endeavour. The proliferation of AI systems imbued with increasingly complex mathematical and data modelling, and machine learning algorithms, are being

integrated in virtually every sector of the economy and society, to support and in many cases undertake more autonomous decisions and actions.

Previously, technologies have often been deployed more like tools, as a pen or paintbrush, but as autonomy and self-learning capabilities increase, robots and intelligent AI systems feel less and less like machines or tools. AI will equip robots and systems with the ability to learn using machine-learning and deep-learning algorithms. They will have the ability to interact and work alongside us or to augment our work. They will increasingly be able to take over functions and roles and, perhaps more significantly, the ability 'to make decisions'.

How much autonomy should AI and robots have to make decisions on our behalf and about us in our life, work and play? How do we ensure they can be trusted, and that they are transparent, reliable, accountable and well designed?

While technological advances hold tremendous promise for mankind, they also pose and raise difficult questions in disparate areas including ethics and morality, bias and discrimination, human rights and dignity, privacy and data protection, data ownership, intellectual property, safety, liability, consumer protection, accountability and transparency, competition law, employment and the future of work and, legal personhood. In a world, that is increasingly connected and where machinebased algorithms use available data to make decisions that affect our lives, how do we ensure these automated decisions are not opaque, appropriate and transparent? And what recourse do we have when these decisions intrude on our rights, freedoms, safety and legitimate interests?

The base tenets of our regulatory systems were created long before the advances and confluence of emergent technologies including AI (artificial intelligence), IoT (Internet of Things), blockchain, cloud, quantum computing, to name a few. With the rise of these technologies we have taken many initiatives to address their consequences by recalibrating and adapting our regulatory frameworks to provide for new opportunities and risks, to confer rights and duties, safety and liability frameworks, and ensure legal certainty for business.

Sector-specific regulation has also been adopted and adapted to address market failures and risks in critical and regulated domains. These changes have often been reactive and piecemeal, with artificial delineation of rights and responsibilities. There have been many unintended consequences. More recently we have begun to learn from past mishaps, and these regulatory adaptations are now more likely to be drafted in technologically neutral way avoiding strict technical definition, especially when the field is still evolving rapidly.

Emerging technologies are rapidly transforming the regulatory landscape. They are providing timely opportunities for fresh approaches in the redesign of our regulatory systems to keep pace with technological changes, now and into the future. AI is currently advancing more rapidly than the process of regulatory recalibration. Unlike the past, there is now a significant difference—we must now take into consideration, machine learning AI systems that have the ability to learn, adapt their performances and 'make decisions' from data and 'life experiences'.

The UN Secretary-General in his June 2020 report, commented that, "The world is at a critical inflection point for technology governance, made more urgent by the ongoing pandemic" [2]. He further emphasized the need to redouble our efforts to better harness digital technologies while mitigating the harm that they may cause.

This paper presented at AI4KM 2020 at the International Joint Conference on Artificial Intelligence - Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI 2020), Yokohama, Japan 7<sup>th</sup> January 2021, provides brief insights on some selected topical developments in our ethical and regulatory systems and, the current debates to address some of the challenges and risks from the use and actions of AI, autonomous and intelligent systems. [1] The paper is partly based on the keynotes, presentations and engagements in Australia [3], Malaysia [4], Zimbabwe [5], Cambodia [6], Sri Lanka [7], Switzerland [8] and Brazil [9]. It extends on the paper published as "The Laws and Regulation of AI and Autonomous Systems" [10].

The paper is organized as follows. Section 2 briefly reviews the state of ethical principles and frameworks. Section 3 looks at the responsibility and liability challenges for damages caused by AI. Section 4 discuss transparency and explainability of AI and section 5 on the debates on Legal personhoods for AI. Section 6 briefly looks at AI and Implications of Employment. Section 7 concludes this paper.

#### 2 Ethical Issues Arising from AI

It is perhaps apt at this juncture, that I pause to reflect on the journey that has brought me to the cross-roads of ethics and regulation of AI. In 2016, I initiated a series of articles on AI for The Australian. The columns commenced with a piece on "Ethics must travel as AI's associate" [11], which was followed by a series of closely related topics including: "How far should AI replace human sense?" [12], "We need plans for when robots are in driver's seat" [13], "Complex algorithms can use a little of that human touch" [14], "AI: Are Musk and Hawking right, or is our future in our hands?" [15], "Do robots and artificial intelligence think about copyright?" [16], "Data frameworks critical for AI success" [17] and, "Who is liable when robots and AI get it wrong?" [18].

These columns led to a series of interviews, panels, and presentations looking at the possible risks that AI poses and the notion of building ethics into machine intelligence. These included an ABC TV News interview on "Neurotechnologies and AI, privacy, agency and identity, and bias", panels to explore the social and ethical concerns of AI and, submission to the consultation on Artificial Intelligence, Australia's Ethics Framework [3].

The many topics canvassed included:

 What happens when AI and algorithmic decision-making leads to someone being disadvantaged or discriminated against? There have been numerous instances where this has happened [19], not necessarily due to the algorithm itself, but because the underlying data reflects an inherent bias, statistical distortion or pattern that becomes obvious when the algorithm is applied to it.

- How do you think traditional business models will be disrupted in the future by AI?
- How will AI disruption of traditional business models impact society?
- What options does the government have to constrain or enable artificial intelligence and what should be its focus?
- What ethical considerations must be taken into account when developing artificial intelligence and what are the priorities?

The conclusions derived from algorithms are probabilistic in nature and may carry inherent biases, which may be replicated, amplified and reinforced. Algorithms are not infallible. As algorithmic complexity and autonomy increases, it becomes imperative to build in checks and balance to protect the legitimate interests of stakeholders [11].

If ethical parameters are programmed into AI, whose ethical and social values are these? This question was foremost in my mind, when I presented on the "Ethical Dimensions of AI & Autonomous Systems" to an audience studying Buddhist ethics [20]. Each society, tradition, cultural group, religion, system and country view ethics and morality through the contextual lenses of their underlying philosophical beliefs. The variations in ethical and social values that underpin our global landscape are challenging and, changes with the passage of time.

In response to the challenges articulated above, a range of stakeholders have produced AI ethical principles and frameworks. When I reviewed AI ethical principles and frameworks produced by public, private, and non-governmental organizations in 2019, there were more than 70 in existence. The number continues to grow. In 2019, jurisdictions including Australia [21] and the EU [22] published their frameworks, adding to the lists of contributors including the OECD Principles on Artificial Intelligence [23], the World Economic Forum AI Governance: A Holistic Approach To Implement Ethics Into AI [24] and the Singapore Model AI Governance Framework [25], to name a few.

An analysis of 84 principles and guidelines by Jobin et al [26] reveals a convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy).

By 2020, a study by Fjeld et al. [27] of 36 principles and guidelines, revealed an extended list around eight key themes: (1) Privacy, (in 97% of documents), (2) Accountability (in 97% of documents), (3) Safety and Security (in 81% of documents), (4) Transparency and Explainability (in 94% of documents), (5) Fairness and Non-discrimination (in 100% of documents), (6) Human Control of Technology (in 69% of documents), (7) Professional Responsibility (in 78% of documents), (8) Promotion of Human Values (in 69% of documents).

The UN Secretary-General in his 2020 report commented that "there are currently over 160 organizational, national and international sets of artificial intelligence ethics and governance principles worldwide" [28] and calls for a common platform to bring these separate initiatives together.

UNESCO was given the mandate by its Member States to develop an international standard-setting instrument on the ethics of artificial intelligence, which is to be submitted to the General Conference in the later part of 2021. The first draft of UNESCO's Recommendation on the Ethics of Artificial Intelligence was released to Member States

in late 2020 [29]. The Recommendation has largely been considered as an inter-disciplinary and multi-stakeholder initiative in light of the proliferation of ethical principles and frameworks.

The Recommendation include many common or shared ethical concepts and values with an extended list around ten key themes: (1) proportionality and do no harm, (2) safety and security, (3) fairness and non-discrimination, (4) sustainability, (5) privacy, (6) human oversight and determination, (7) transparency and explainability, (8) responsibility and accountability, (9) awareness and literacy and (10) multi-stakeholder and adaptive governance and collaboration.

The debates have matured significantly since 2017, beyond the 'what' of ethical principles to more of the 'how', with detailed guidelines on how such principles can be operationalised in the design and implementation to minimise risks and negative outcomes. But the challenge has always been putting principles into practice and creating accountability mechanisms.

There is a growing consensus that the time has arrived to move from principles and to operationalize on the ethical practice on AI [30]. Many of the proponents for regulatory intervention have argued that abstract high-level AI principles lack the specificity to be used in practice and require legal enforcement mechanisms that are more robust to provide redress when things go wrong. With the growing lists of AI related incidents, there is a general distrust that AI developers could self-regulate effectively.

There is also a growing awareness that principles can play a useful base from which to develop professional ethics, standards, and AI regulatory systems across the globe. But how do we translate and embrace these articulated values, principles and actions to guide Nation States in the formulation of their regulatory systems, policies or other legal instruments regarding AI?

As stated by Fjeld et al., the impact of a set of principles is "likely to depend on how it is embedded in a larger governance ecosystem, including for instance relevant policies (e.g. AI national plans), laws, regulations, but also professional practices and everyday routines" [31]. That view also resonated with those of UNESCO. UNESCO has advocated for Member States to put in place policy actions and oversight mechanisms to operationalize the values and principles in the Recommendation.

Due to the challenges in enforcing ethical principles or frameworks, we are seeing greater regulatory impetus and focus to complement the gaps to improve the public's trust. We are seeing growing awareness that our existing regulatory frameworks are not evolving fast enough to keep pace with the rapid progress in AI. Recently, UNESCO and the EU Parliament have set the regulatory train in motion.

One of the objectives of the UNESCO Recommendation is to provide a universal framework of values, principles and actions to guide Member States in the formulation of their legislation, policies or other instruments regarding AI: "Member States should develop, review and adapt, as appropriate, regulatory and legal frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. Member States should introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and behaviour of AI systems" [32].

UNESCO has strongly advocated that AI cannot be a no law zone: "There are some legislative vacuums around the industry which needs to be filled fast. The first step is to agree on exactly which values need to be enshrined, and which rules need to be enforced. Many frameworks and guidelines exist, but they are implemented unevenly, and none are truly global. AI is global, which is why we need a global instrument to regulate it" [34].

In October 2020, the European Parliament adopts 3 resolutions to regulate AI, setting the pace as a global leader in AI regulation. The resolutions cover the ethical and legal obligations surrounding AI, civil liability setting fines of up to 2 million euros for damage caused by AI; and intellectual property rights [35]. In response, the European Commission has indicated that it will publish draft legislation in 2021 addressing AI by obliging high-risk AI systems to meet mandatory requirements related to their trust-worthiness.

2021 will prove to be an interesting year for AI regulatory developments. However, what will unfold, time will tell. Some of these ethical principles or frameworks may well be adopted alongside or incorporated in legislation.

#### 3 Responsibility and Liability for damages caused by AI

How should regulators manage the complexity and challenges arising from the design, development and deployment of robots and autonomous systems? What legal and social responsibilities should we give to algorithms shielded behind statistically data-derived 'impartiality'? Who is liable when robots and AI get it wrong?

There is much debate as to who amongst the various players and actors across the design, development and deployment lifecycle of AI and autonomous systems should be responsible and liable to account for any damages that might be caused. Would autonomy and self-learning capabilities alter the chain of responsibility of the producer or developer as the "AI-driven or otherwise automated machine which, after consideration of certain data, has taken an autonomous decision and caused harm to a human's life, health or property" [36]?

Or has "inserting a layer of inscrutable, unintuitive, and statistically-derived code in between a human decisionmaker and the consequences of that decision, AI disrupts our typical understanding of responsibility for choices gone wrong"? [37] Or should the producer or programmer foresee the potential loss or damage even when it may be difficult to anticipate—particularly in unusual circumstances, the actions of an autonomous system? These questions will become more critical as more and more autonomous decisions are made by AI systems.

One of the more advanced regulatory developments in AI is in the trialling of autonomous vehicles [38] and in the regulatory frameworks for drones [39].

The rapid adoption of AI and autonomous systems into more diverse areas of our lives—from business, education, healthcare and communication through to infrastructure, logistics, defence, entertainment and agriculture—means that any laws involving liability will need to consider a broad range of contexts and possibilities.

We are moving rapidly towards a world where autonomous and intelligent AI systems are connected, embedded and integrated in complex environments, and with "the plurality of actors involved, it can be difficult to assess where a potential damage originates and which person is liable for it. Due to the complexity of these technologies, it can be very difficult for victims to identify the liable person and prove all necessary conditions for a successful claim, as required under national law" [40]. That view is also reflected in the more recent European Parliament resolution with recommendations to the Commission on a civil liability regime for artificial intelligence [41]. The burden of proof in a tort fault-based liability system in some countries could significantly increase the costs of litigation.

We will need to establish specific protections for potential victims of AI-related incidents to give consumers confidence that they will have legal recourse if something goes wrong.

One of the proposals being debated is for the creation of a mandatory insurance scheme to ensure that victims of incidents involving robots and intelligent AI systems have access to adequate compensation. This might be similar to the mandatory comprehensive insurance that owners need to purchase before being able to register a motor vehicle [42]. The EU Parliament has recently also proposed for deployers of high-risk AI to have mandatory liability insurance ( $\epsilon$ 10m in the event of death and physical harm and  $\epsilon$ 2m for damage to property) [41].

Another approach is for the creation of strict liability rules to compensate victims for potential harm caused by AI and autonomous systems along the lines of current product liability laws in the EU and Australia. Strict liability rules would ensure that the victim is compensated regardless of fault. The EU Parliament has proposed that deployers of AI designated as "high-risk" would be strictly liable for any damage caused by it. But who amongst the various players and actors should be strictly liable?

Whether the existing mixture of fault-based and strict liability regimes are appropriate is also subject to much debate.

Introducing a robust regulatory framework with relevant input from industry, policymakers and government would create greater incentive for AI developers and manufacturers to reduce their exposure by building in additional safeguards to minimise the potential risks to humanity.

#### 4 Transparency and Explainability of AI

Algorithms are increasingly being used to analyse information and define or predict outcomes with the aid of AI. These AI systems may be embedded in devices and systems and deployed across many industries and increasingly in critical domains, often without the knowledge and consent of the user. Should humans be informed that they are interacting with AI, on the purposes of the AI, and on the data used for the training and evaluation?

To ensure that AI based systems perform as intended, the quality, accuracy and relevance of data are essential. Any data bias, error or statistical distortion will be learned

and amplified. In situations involving machine learning—where algorithms and decision rules are trained using data to recognize patterns and to learn to make future decisions based on these observations, regulators and consumers may not easily discern the properties of these algorithms. These algorithms are able to train systems to perform certain tasks at levels that may exceed human ability and raise many challenging questions including calls for greater algorithmic transparency to minimise the risk of bias, discrimination, unfairness, error and to protect consumer interests.

Over the last few years legislators have started to respond to the challenge. In the EU, Article 22 of the General Data Protection Regulation (GDPR) [43] gives individuals the right not to be subject to a decision based solely on automated decision-making (no human involvement in the decision process), except in certain situations including explicit consent and necessity for the performance of or entering into a contract. The GDPR applies only to automated decision-making involving personal data.

In the public sector, AI systems are increasingly being adopted by governments to improve and reform public service processes. In many situations, stakeholders and users of AI will expect reasons to be given for transparency and accountability of government decisions which are important elements for the proper functioning of public administration. It is currently unclear how our regulatory frameworks would adjust to providing a meaningful review by our courts of decisions undertaken by autonomous AI systems, or in what circumstances a sub-delegation by a nominated decision-maker to an autonomous AI systems would be lawful. We may need to develop new principles and standards and "to identify directions for thinking about how administrative law should respond ... that makes sense from both a legal and a technical point of view. [44].

As machine learning evolves, AI models [45] often become even more complex, to the point where it may be difficult to articulate and understand their inner workings—even to people who created them. This raises many questions: what types of explanation are suitable and useful to the audience? [46] How and why does the model perform the way it does? How comprehensive does the explanation need to be—is an understanding on how the algorithmic decision was reached required, or should the explanation be adapted in a manner which is useful to a non-technical audience?

In the EU, the GDPR explicitly provides a data subject with the following rights:

- a) rights to be provided and to access information about the automated decision-making; [47]
- b) rights to obtain human intervention and to contest the decision made solely by automated decision-making algorithm; [48] and
- c) places explicit onus on the algorithmic provider to provide "meaningful information about the logic involved" in algorithmic decision, the "significance" and the "envisaged consequences" of the algorithmic processing [49].

But how would these rights operate and be enforced in practice? With recent and more complex non-linear black-box AI models, it can be difficult to provide meaningful explanations, largely due to the statistical and probabilistic character of machine learning and the current limitations of some AI models—raising concerns including accountability, explainability, interpretability, transparency, and human control.

What expertise and competencies would be required from a data subject to take advantage of the rights or for the algorithmic provider to provide the above rights?

"In addition, access to the algorithm and the data could be impossible without the cooperation of the potentially liable party. In practice, victims may thus not be able to make a liability claim. In addition, it would be unclear, how to demonstrate the fault of an AI acting autonomously, or what would be considered the fault of a person relying on the use of AI" [50].

This opacity will also make it difficult to verify whether decisions made with the involvement of AI are fair and unbiased, whether there are possible breaches of laws, and whether they will hamper the effective access to the traditional evidence necessary to establish a successful liability action and to claim compensation.

Should organisations consider and ensure that specific types of explanation be provided for their proposed AI system to meet the requisite needs of the audience before starting the design process? Should the design and development methodologies adopted have the flexibility to embrace new tools and explanation frameworks, ensuring ongoing improvements in transparency and explainability in parallel with advancement in the state of the art of the technology throughout the lifecycle of the AI system?

While rapid development methodologies may have been adopted by the IT Industry, embedding transparency and explainability into AI system design requires more extensive planning and oversight, and requiring input and knowledge from a wider mix of multi-disciplinary skills and expertise.

New tools and better explanation frameworks need to be developed to instill the desired human values and to reconcile the current tensions and trade-off between accuracy, cost and explainability of AI models. Developing such tools and frameworks is far from trivial, warranting further research and funding.

#### 5 Legal personhoods for AI

Historically, our regulatory systems have granted rights and legal personhood to slaves, women, children, corporations and more recently to landscape and nature. Two of India's rivers, the Ganga and the Yamuna, have been granted legal status. In New Zealand legislation was enacted to grant legal personhoods to the Whanganui river, Mount Taranaki and the Te Urewera protected area. Previously, corporations were the only nonhuman entities recognised by the law as legal persons.

"To be a legal person is to be the subject of rights and duties" [51]. Granting legal personality [52] to AI and robots will entail complex legal considerations and is not a simple case of equating them to corporations.

Who foots the bill when a robot or an intelligent AI system makes a mistake, causes an accident or damage, or becomes corrupted? The manufacturer, the developer, the person controlling it, or the robot itself? Or is it a matter of allocating and apportioning risk and liability?

As autonomic and self-learning capabilities increase, robots and intelligent AI systems will feel less and less like machines and tools. Self-learning capabilities for AI have added complexity to the equation. Will granting 'electronic rights' to robots assist

with some of these questions? Will human actors use robots to shield themselves from liability or shift any potential liabilities from the developers to the robots? Or will the spectrum, allocation and apportionment of responsibility keep step with the evolution of self-learning robots and intelligent AI systems? Regulators around the world are wrestling with these questions.

The EU is leading the way on these issues. In 2017 the European Parliament, in an unprecedented show of support, adopted a resolution on Civil Law Rules on Robotics [53] by 396 votes to 123. One of its key recommendations was to call on the European Commission to explore, analyse and consider "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 responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions" [54].

The EU resolution generated considerable debate and controversy, because it calls for sophisticated autonomous robots to be given specific legal status as electronic persons. The arguments from both sides are complex and require fundamental shifts in legal theory and reasoning.

In an open letter, experts in robotics and artificial intelligence have cautioned the European Commission that plans to grant robots legal status are inappropriate and "non-pragmatic" [55].

The European Group on Ethics in Science and New Technologies, in its Statement on Artificial Intelligence, Robotics and Autonomous Systems, advocated that the concept of legal personhood is the ability and willingness to take and attribute moral responsibility. "Moral responsibility is here construed in the broad sense in which it may refer to several aspects of human agency, e.g. causality, accountability (obligation to provide an account), liability (obligation to compensate damages), reactive attitudes such as praise and blame (appropriateness of a range of moral emotions), and duties associated with social roles. Moral responsibility, in whatever sense, cannot be allocated or shifted to 'autonomous' technology" [56].

In 2020, the EU Commission presented its "White Paper on Artificial Intelligence—A European approach to excellence and trust for regulation of artificial intelligence (AI)" [57] and a number of other documents including a "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics" [58] for comments. The White Paper is non-committal on the question of endowing robots with specific legal status as electronic persons. It proposes a risk-based approach to create an 'ecosystem of trust' as one of the key elements of a future regulatory framework for AI in Europe, so that the regulatory burden is not excessively prescriptive or disproportionate.

I concur with the conclusions reached by Bryson et al [59] that the case for electronic personhood is weak. With the current capabilities and state-of-the-art of AI systems, it is essential that human stays 'in the loop'. The negatives outweigh the benefits in the current debate on shifting legal and moral responsibility to AI systems—at least for the foreseeable future. That view is consistent with those reached by UNESCO: "when developing regulatory frameworks, Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or

legal persons and that AI systems should not be given legal personality themselves" [60].

In October 2020, the EU Parliament reversed its earlier resolution and makes it clear that it would not be appropriate to grant legal personhood to AI [61].

As evidenced by the historical debates on the status of slaves, women, corporations and, more recently landscape and nature, the question of granting legal personality to autonomous robots will not be resolved any time soon. There is no simple answer to the question of legal personhood, and one size will not fit all.

Should legal personhood for robots or autonomous systems eventuate in the future, any right invoked on behalf of robots, or obligation enforced against them, will require new approaches and significant recalibration of our regulatory systems. Legal personhood could potentially allow autonomous robots to own their creations, as well as being open to liability for problems or negative outcomes associated with their actions.

### 6 AI and Implications of Employment

Over the past few years we have been inundated with predictions that robots and automation will devastate the workplace, replacing many job functions within the next 10 to 15 years. We have already seen huge shifts in manufacturing, mining, agriculture, administration and logistics, where a wide range of manual and repetitive tasks have been automated. More recently, cognitive tasks and data analyses are increasingly being performed by AI and machines.

Historically, new technologies have always affected the structure of the labour market, leading to a significant impact on employment, especially lower skilled and manual jobs. But now the pace and spread of autonomous and intelligent technologies are outperforming humans in many tasks and radically challenging the base tenets of our labour markets and laws. These developments have raised many questions.

Where are the policies, strategies and regulatory frameworks to transition workers in the jobs that will be the most transformed, or those that will disappear altogether due to automation, robotics and AI?

Our current labour and employment laws, such as sick leave, hours of work, tax, minimum wage and overtime pay requirements, were not designed for robots. What is the legal relationship of robots to human employees in the workplace? In relation to workplace safety— what liabilities should apply if a robot harms a human co-worker? Would the 'employer' of the robot be vicariously liable? What is the performance management and control plan for work previously undertaken by human employees working under a collective bargaining agreement, now performed or co-performed with AI or robots? How would data protection and privacy regulations apply to personal information collected and consumed by robots? Who would be responsible for cyber security and the criminal use of robots or AI? [62]

Are there statutory protection and job security for humans displaced by automation and robots? Should we tax robot owners to pay for training for workers who are displaced by automation, or should there be a universal minimum basic income for people displaced? Should we have social plans, such as exist in Germany and France, if restructuring through automation disadvantages employees?

There are many divergent views on all these questions. All are being hotly debated. Governments, policy makers, institutions and employers all have important roles to play in the development of digital skills, in the monitoring of long-term job trends, and in the creation of policies to assist workers and organisations adapt to an automated future. If these issues are not addressed early and proactively, they may worsen the digital divide and increase inequalities between countries and people.

ICT professionals are also being impacted as smart algorithms and other autonomous technologies supplement software programming, data analysis and technical support roles. With AI and machine learning developing at an exponential rate, what does the future look like?

#### 6.1 Case study - line between human and robo advisers in financial services

FinTech (financial technology) start-ups are emerging to challenge the roles of banks and traditional financial institutions. FinTechs are rapidly transforming and disrupting the marketplace by providing 'robo-advice' using highly sophisticated algorithms operating on mobile and web-based environments. The technology is called robotic process automation (RPA) and is becoming widespread in business, and particularly in financial institutions. Robo-advice or automated advice is the provision of automated financial product advice using algorithms and technology and without the direct involvement of a human adviser [63].

Robo-advice and AI capabilities have the potential to increase competition and lower prices for consumers in the financial advice and financial services industries by radically reshaping the customer experience. They are designed, modelled and programmed by human actors. Often they operate behind the scenes 24/7 assisting the people who interact with consumers. There are considerable tasks and risks involved in writing algorithms to accurately portray the full offerings and complexity of financial products.

In 2017 Australia, after a number of scandals, introduced professional standards legislation for human financial advisers [64]. These regulations set higher competence and ethical standards, including requirements for relevant first or higher degrees, continuing professional development requirements and compliance with a code of ethics. The initiatives were introduced into a profession already under pressure from the robo environment.

Because robo-advice is designed, modelled and programmed by human actors, should these requirements also apply to robo-advice? Should regulators also hold ICT developers and providers of robots and autonomous systems to the same standards demanded from human financial advisers? What should be the background, skills and competencies of these designers and ICT developers?

Depending on the size and governance framework of an organisation, various players and actors could be involved in a collaborative venture in the development, deployment and lifecycle of AI systems. These might include the developer, the product manager, senior management, the service provider, the distributor and the person who uses the AI or autonomous system. Their domain expertise could be in computer science, or

mathematics or statistics, or they might be an interdisciplinary group composed of financial advisers, economists, social scientists or lawyers.

In 2016 the Australian regulator laid down sectoral guidelines [65] for monitoring and testing algorithms deployed in robo-advice. The regulatory guidance requires businesses offering robo-advice to have people within the business who understand the "rationale, risk and rules" used by the algorithms and have the skills to review the resulting robo-advice. What should be the competencies and skills of the humans undertaking the role?

The EU General Data Protection Regulation (GDPR) [66] went further, by placing an explicit onus on the algorithmic provider to provide "meaningful information about the logic involved" [67]. In addition, GDPR provides an individual with explicit rights including the rights to obtain human intervention, to express their point of view and to contest the decision made solely by automated systems [68] that has legal or similarly significant impact. GDPR applies only when AI uses personal data within the scope of the legislation.

Revealing the logic behind an algorithm may potentially risk and disclose commercially sensitive information and trade secrets used by the AI model and on how the system works.

The deployment of robo-advice raises many new, interesting and challenging questions for regulators accustomed only to assessing and regulating human players and actors.

#### 7 Conclusion

This paper raises some of the major topical issues and debates relating to ethics of AI, AI liability, transparency and meaningful AI explanation, aspects of data protection and privacy, legal personhood, job transition and employment law.

In the wake of the 2020 "black lives matter" protests, a number of technology companies have announced limitations on plans to sell facial recognition technology. There have also been renewed calls for a moratorium on certain uses of facial recognition technology that has legal or significant effects on individuals until appropriate legal framework has been established [69].

The need to address AI and autonomous system challenges has increased in urgency as the adverse potential impact could be significant in specific critical domains. If not appropriately addressed, human trust will suffer, impacting on adoption and oversight and in some cases posing significant risks to humanity and societal values.

From this brief exploration, it is clear that the values and issues outlined in the paper will benefit from much broader debate, research and consultation. There are no definitive answers to some of the questions raised—as for many, it is a matter of perspective. I trust that this paper will embark you on your own journey as to what our future regulatory systems should encapsulate. Different AI applications create and pose different benefits, risks and issues. The solutions that might be adopted in the days ahead, will potentially challenge our traditional beliefs and systems for years to come. We are fac-

ing a major disruptive shift which is capable of dislodging some of our legal assumptions and may require significant rethink of some of our long-established legal principles—as we must now take into consideration, machine learning AI systems that have the ability to learn, adapt and 'make decisions' from data and 'life experiences'.

Technologists and AI developers understand better than most in relation to the trends and trajectories of emergent technologies and their potential impact on the economic, safety and social constructs of the workplace and society. Is it incumbent on them to raise these issues and ensure they are widely debated, so that appropriate and intelligent decisions can be made for the changes, risks and challenges ahead? Technologists and AI developers are well placed to address some of the risks and challenges during the design and lifecycle of AI-enabled systems. It would be beneficial to society for ICT professionals to assist government, legislators, regulators and policy formulators with their unique understanding of the strengths and limitations of the technology and its effects.

Historically, our regulatory adaptations have been conservative and patchworked in their ability to keep pace with technological changes. Perhaps the drastic disruptions that COVID-19 has caused in our work, life and play beyond the normal will provide sufficient impetus and tenacity to consider and re-think on how our laws and regulatory systems should recalibrate with AI and autonomous systems, now and into the future.

- This paper is for general reference purposes only. It does not constitute legal or professional
  advice. It is general comment only. Before making any decision or taking any action you
  should consult your legal or professional advisers to ascertain how the regulatory system
  applies to your particular circumstances in your jurisdiction
- United Nation, Report of the Secretary-General, Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation, www.un.org/en/content/digital-cooperation-roadmap/, June 2020, p 3, www.un.org/en/content/digital-cooperation-roadmap/, accessed January 2021
- 3. Neurotechnologies and AI, privacy, agency and identity, and bias, ABC TV News 24 November 2017; Re-engineering industries with Artificial Intelligence & the social contract. The intended outcome, 26th International Joint Conference on Artificial Intelligence, Melbourne August 2017, https://www.acs.org.au/insightsandpublications/media-releases/00000121.html; Automation and The Nature of Work, ACS Canberra Conference, August 2017; Robo-advice & FinTech: More Transparent, Honest & Reliable than Human Actors? Chair and Panel, Sydney University Business School, November 2018; Future AI Forum submission to the consultation, Artificial Intelligence, Australia's Ethics Framework, May 2019, https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/consultation/view\_respondent?\_b\_index=60&uuId=309030657; Ethical Dimensions of AI & Autonomous Systems, Nan Tien Institute, Wollongong, Australia, August 24 2019
- International ICT Infrastructure & Digital Economy Conference Sarawak (IDECS) 2017 Kuching, Sarawak, Malaysia, April 2017; Presentation to Global Conference on Computing Ethics, Kuala Lumpur, Malaysia, August 2012.
- 5. Artificial Intelligence, the Good, and the Ugly, Victoria Falls, Zimbabwe, November 2017
- AI and Employment Law Implications, LAWASIA 2018, Siem Reap, Cambodia, November 2018

- The Ethics & Regulation of Artificial Intelligence, Keynote to NITC 2019 "Embrace Digital" Colombo, Sri Lanka, 9th October 2019
- 8. Duty of Care and Ethics on Digital Technologies, Internet Governance Forum (IGF), Geneva, December 2017; How do we maximise the benefits of Innovative 4.0 technologies, without unnecessary risks and consequences?, World Summit on the Information Society (WSIS), Geneva, May 2019; AI: Complementing Codes of Ethics with Law and Regulation, Session 131 "Living the standard how can the Information and Knowledge Society live to an ethical and FAIR Standard", World Summit on the Information Society (WSIS), Geneva, July 2020
- Ethics & Regulation of Artificial Intelligence, SECOMU 2020 "Artificially Human or Humanly Artificial? Challenges for Society 5.0", Brazil 18th November 2020
- Wong A. (2020) The Laws and Regulation of AI and Autonomous Systems. In: Strous L., Johnson R., Grier D.A., Swade D. (eds) Unimagined Futures – ICT Opportunities and Challenges. IFIP Advances in Information and Communication Technology, vol 555. Springer, Cham. https://doi.org/10.1007/978-3-030-64246-4\_4
- 11. "Ethics must travel as AI's associate", The Australian, September 13, 2016
- 12. "How far should AI replace human sense?", The Australian, September 27, 2016
- 13. "We need plans for when robots are in driver's seat", The Australian, November 22, 2016
- 14. "Complex algorithms can use a little of that human touch", The Australian, January 24, 2017
- 15. "AI: Are Musk and Hawking right, or is our future in our hands?", The Australian, August 23, 2017
- "Do robots and artificial intelligence think about copyright?", The Australian, September 5, 2017
- 17. "Data frameworks critical for AI success", The Australian, October 4, 2017
- 18. "Who is liable when robots and AI get it wrong?", The Australian, September 19, 2017
- Refer to examples in the AI Incident database https://incidentdatabase.ai/summaries/incidents
- Ethical Dimensions of AI & Autonomous Systems, Nan Tien Institute, Wollongong, Australia, August 24 2019
- Australian AI Ethics Framework (2019). https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework, last accessed 2020/6/6
- European Commission: Ethics guidelines for trustworthy AI (2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, last accessed 2020/6/6
- 23. OECD, OECD Principles on Artificial Intelligence (22 May 2019), https://www.oecd.org/going-digital/ai/principles/, last accessed 2020/6/20
- World Economic Forum: AI Governance: A Holistic Approach to Implement Ethics into AI, https://www.weforum.org/whitepapers/ai-governance-a-holistic-approach-to-implementethics-into-ai, last accessed 2020/6/20
- Singapore Model AI Governance Framework, https://www.pdpc.gov.sg/-/me-dia/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf, last accessed 2020/6/20
- Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. Nat Mach Intell 1, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
- 27. Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research

- Publication No. 2020-1, https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482
- United Nation, Report of the Secretary-General, Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation, www.un.org/en/content/digital-cooperation-roadmap/, June 2020 p 18, www.un.org/en/content/digital-cooperation-roadmap/ accessed January 2021
- 29. The first draft of the recommendation submitted to Member States proposes options for action to Member States and other stakeholders and is accompanied by concrete implementation guidelines. The first draft of the AI Ethics Recommendation is available at https://unesdoc.unesco.org/ark:/48223/pf0000373434
- 30. See also the opinion of the High Level Panel Follow-up Roundtable 3C Artificial Intelligence 1st Session, www.un.org/en/pdfs/HLP%20Followup%20Roundtable%203C%20Artificial%20Intelligence%20-%201st%20Session%20Summary.pdf
- Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482, page 5
- 32. UNESCO first draft of the AI Ethics Recommendation, Resolution 68, https://unesdoc.unesco.org/ark:/48223/pf0000373434
- 33. European Parliament three Resolutions on the ethical and legal aspects of Artificial Intelligence software systems ("AI"): Resolution 2020/2012(INL) on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and related Technologies (the "AI Ethical Aspects Resolution"), Resolution 2020/2014(INL) on a Civil Liability Regime for Artificial Intelligence (the "Civil Liability Resolution"), and Resolution 2020/2015(INI) on Intellectual Property Rights for the development of Artificial Intelligence Technologies (the "IPR for AI Resolution")
- 34. Elaboration of a Recommendation on ethics of artificial intelligence (unesco.org), https://en.unesco.org/artificial-intelligence/ethics
- 35. EU Parliament three Resolutions on the ethical and legal aspects of Artificial Intelligence software systems ("AI"): Resolution 2020/2012(INL) on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and related Technologies (the "AI Ethical Aspects Resolution"), Resolution 2020/2014(INL) on a Civil Liability Regime for Artificial Intelligence (the "Civil Liability Resolution"), and Resolution 2020/2015(INI) on Intellectual Property Rights for the development of Artificial Intelligence Technologies (the "IPR for AI Resolution").
- 36. The World Economic Forum; White Paper on AI Governance A Holistic Approach to Implement Ethics into AI, p. 6. Geneva, Switzerland (2019), https://www.weforum.org/white-papers/ai-governance-a-holistic-approach-to-implement-ethics-into-ai, last accessed 2020/6/9
- Selbst, Andrew D.: Negligence and AI's Human Users. In: Public Law & Legal Theory Research Paper No. 20-01, p 1. UCLA School of Law (2018)
- 38. For a brief rundown of the regulatory frameworks and developments in selected countries refer to the Australian National Transport Commission 2020, Review of 'Guidelines for trials of automated vehicles in Australia': Discussion paper, NTC, Melbourne, pp. 16-18, https://www.ntc.gov.au/sites/default/files/assets/files/NTC%20Discussion%20Paper%20-%20Review%20of%20guidelines%20for%20trials%20of%20automated%20vehicles%20in%20Australia.pdf, last accessed 2020/6/6. For examples of Australian legislation

- refer to: Motor Vehicles (Trials of Automotive Technologies) Amendment Act 2016 (SA), Transport Legislation Amendment (Automated Vehicle Trials and Innovation) Act 2017 (NSW), Road Safety Amendment (Automated Vehicles) Act 2018 (Vic)
- 39. For the new European Union drone rules refer to: https://www.easa.europa.eu/do-mains/civil-drones-rpas/drones-regulatory-framework-background. For the Australia drone rules refer to: https://www.casa.gov.au/knowyourdrone/drone-rules and the Civil Aviation Safety Amendment (Remotely Piloted Aircraft and Model Aircraft—Registration and Accreditation) Regulations 2019
- 40. European Commission: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM (2020) 64 (Feb. 19, 2020), p 14, https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-thingsand-robotics\_en, last accessed 2020/6/
- 41. European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence: www.europarl.europa.eu/doceo/document/TA-9-2020-0276\_EN.pdf
- 42. Australian National Transport Commission 2020, Review of 'Guidelines for trials of automated vehicles in Australia': Discussion paper, NTC, Melbourne, pp. 26-27, https://www.ntc.gov.au/sites/default/files/assets/files/NTC% 20Discussion% 20Paper% 20% 20Review% 20of% 20guidelines% 20for% 20trials% 20of% 20automated% 20vehicles% 20in% 20Australia.pdf, last accessed 2020/6/6
- 43. General Data Protection Regulation (GDPR) art.22; Recital 71; see also Article 29 Data Protection Working Party, 2018a, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 17/EN WP251rev.01, p.19. http://ec.europa.eu/newsroom/article29/item-detail.cfm?item\_id=612053, last accessed 2020/6/4
- Cobbe, Jennifer.: Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making. Legal Studies, p 3 (2019)
- 45. For the interpretability characteristics of various AI models, refer to ICO and Alan Turing Institute: Guidance on explaining decisions made with AI (2020), annexe 2, https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf, last accessed 2020/6/6
- 46. For the types of explanation that an organisation may provide, refer to ICO and Alan Turing Institute: Guidance on explaining decisions made with AI (2020), p. 20, https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf, last accessed 2020/6/6
- 47. General Data Protection Regulation (GDPR) art.15
- 48. General Data Protection Regulation (GDPR) art.22
- 49. General Data Protection Regulation (GDPR) arts.13-14
- 50. European Commission: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM (2020) 64 (Feb. 19, 2020), p 15, https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-thingsand-robotics\_en, last accessed 2020/6/9
- 51. Smith, B.: Legal personality. Yale Law J 37(3), 283-299 (1928), p 283
- 52. For a discussion on the concept and expression "legal personality" refer to Bryson, J. J., Diamantis, M. E., Grant, T.D.: Of, for, and by the people: the legal lacuna of synthetic persons. Artificial Intelligence and Law. 25(3) (2017), p. 277
- 53. European Parliament: European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), https://eur-

- lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017IP0051, last accessed 2020/6/9
- 54. Ibid paragraph 59(f)
- 55. Refer http://www.robotics-openletter.eu/, last accessed 2020/6/9
- European Group on Ethics in Science and New Technologies: Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, p 10. European Commission, Brussels (2018), http://ec.europa.eu/research/ege/pdf/ege\_ai\_statement\_2018.pdf, last accessed 2020/6/9
- 57. European Commission: White Paper on Artificial Intelligence A European approach to excellence and trust, COM(2020) 65 (Feb. 19, 2020), https://ec.europa.eu/info/sites/info/files/commissionwhite-paper-artificial-intelligence-feb2020\_en.pdf, last accessed 2020/6/9
- 58. European Commission: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM (2020) 64 (Feb. 19, 2020), https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-thing-sand-robotics\_en, last accessed 2020/6/9
- 59. Bryson, J. J., Diamantis, M. E., Grant, T.D.: Of, for, and by the people: the legal lacuna of synthetic persons. Artificial Intelligence and Law. 25(3) (2017), pp. 273-291
- 60. UNESCO first draft of the AI Ethics Recommendation, Resolution 68, available at https://unesdoc.unesco.org/ark:/48223/pf0000373434
- 61. Resolution 2020/2014(INL) on a Civil Liability Regime for Artificial Intelligence (the "Civil Liability Resolution"), "any required changes in the existing legal framework should start with the clarification that AI-systems have neither legal personality nor human conscience, and that their sole task is to serve humanity" (Civil Liability Resolution, Annex, (6)); European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (2020/2015(INI)), "that it would not be appropriate to seek to impart legal personality to AI technologies and points out the negative impact of such a possibility on incentives for human creators", para 13
- 62. This section on employment implications is based on the presentation to LAWASIA 2018 at Siem Reap, Cambodia 4 November 2018: Artificial Intelligence & Employment Law Implications
- 63. Definition from the Australian Securities & Investments Commission: Regulatory Guide 255 - Providing digital financial product advice to retail client, https://asic.gov.au/regulatory-resources/find-a-document/regulatory-guides/rg-255-providing-digital-financial-product-advice-to-retail-clients/, last accessed 2020/6/6
- 64. Corporations Amendment (Professional Standards of Financial Advisers) Act 2017
- Australian Securities & Investments Commission: Regulatory Guide 255 Providing digital financial product advice to retail client, https://asic.gov.au/regulatory-resources/find-a-document/regulatory-guides/rg-255-providing-digital-financial-product-advice-to-retail-clients/, last accessed 2020/6/6
- 66. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1) [GDPR].
- 67. Ibid art. 15(1)(h)
- 68. Ibid art. 22(3)
- 69. Australian Human Rights Commission: Discussion Paper on Human Rights and Technology (2019), p 104, https://humanrights.gov.au/our-work/rights-and-freedoms/publications/hu-

man-rights-and-technology-discussion-paper-2019, last accessed 2020/6/20; For a US perspective, refer to Flicker, Kirsten: The Prison of Convenience, The Need for National Regulation of Biometric Technology in Sports Venues In: 30 Fordham Intell. Prop. Media & Ent.L.J. 985 (2020), p 1015, https://ir.lawnet.fordham.edu/iplj/vol30/iss3/7/, last accessed 2020/6/20